PARAMETER ESTIMATION – 5

Practical implementation of parameter estimation
Evaluation of the results

Anna Ibolya Pózna

University of Pannonia
Faculty of Information Technology
Department of Electrical Engineering and Information Systems

pozna.anna@virt.uni-pannon.hu

November 16, 2020

Contents
Lectures and tutorials

- Basic notions, Elements of random variables and mathematical statistics
- The properties of the estimates, Linear regression
- Stochastic processes, Discrete time stochastic dynamic models
- Least squares (LS) estimation by minimizing the prediction error, The properties of the LS estimation
- Special methods for LS estimation of dynamic model parameters: Instrumental variable (IV) method, Parameter estimation of dynamic nonlinear models
- Practical implementation of parameter estimation: Data checking and preparation, Evaluation of the results of parameter estimation

Lecture overview

1. Statistical properties of the dynamic LS estimate
   - Conditions for asymptotic unbiasedness
2. Preparing and checking measurement data
3. Experiment design
   - Sufficient exitation
   - PRBS test signal
4. Evaluating the quality of the estimate
   - Analyzing the residuals/prediction errors
   - Analysing the covariances of the estimates
   - Nonlinear case - an example

Overview

Recall Dynamic LS estimate

Dynamic predictive model: $y(k) = p_0^\top \cdot \varphi(k) + \nu_0(k)$

Important (LS estimate)

$$\hat{p}_{LS} = \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\varphi^\top(k)\right]^{-1} \cdot \frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot y(k)$$

Important (Estimation error)

$$\hat{p}_{LS}(N) = p_0 + [R(N)]^{-1}\,\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\nu_0(k)$$

The estimation error is the second term in the above equation.

(Asymptotic unbiasedness)

is the property of the estimate when the sample size is growing.

Recall, that a dynamic predictive model can be expressed with the **nominal** model (with nominal parameter vector p_0^T and the regressor $\varphi(k)$), and a noise term $\nu_0(k)$. The noise can be measurement or modelling noise.
$y(k) = p_0^\top \cdot \varphi(k) + \nu_0(k)$

- The LS estimate of p can be written as before:
  $\hat{p}_{LS} = \left[\frac{1}{N}\sum_{k=1}^{N} \varphi(k) \cdot \varphi^\top(k)\right]^{-1} \cdot \frac{1}{N}\sum_{k=1}^{N} \varphi(k) \cdot y(k)$
- We can denote the term $\frac{1}{N}\sum_{k=1}^{N} \varphi(k) \cdot \varphi^\top(k)$ by $R(N)$.
- Substituting $R(N)$ and $y(k)$ to the equation of the LS estimate, it can be be written as
  $\hat{p}_{LS}(N) = [R(N)]^{-1} \frac{1}{N}\sum_{k=1}^{N} \varphi(k) \left[\varphi(k)^\top \cdot p_0 + \nu_0(k)\right]$
- Resolving the braces, the first term on the right side is
  $[R(N)]^{-1} \frac{1}{N}\sum_{k=1}^{N} \varphi(k) \cdot \varphi(k)^\top \cdot p_0 = [R(N)]^{-1} R(N) \cdot p_0 = p_0$
- The second term is $[R(N)]^{-1} \frac{1}{N}\sum_{k)=1}^{N} \varphi(k) \cdot \nu_0(k)$
- Therefore the LS estimate can be written in the following form $\hat{p}_{LS}(N) = p_0 + [R(N)]^{-1} \frac{1}{N}\sum_{k=1}^{N} \varphi(k) \cdot \nu_0(k)$. It can be seen that it is composed of the true value of p (p_0) and the estimation error .

In case of dynamic models, the unbiasedness of the estimate can be examined from an asymptotic point of view.

Recall Stochastic properties

$$y(k) = p_0^\top \cdot \varphi(k) + \nu_0(k)$$

When the $\nu_0(k)$ error is small compared to the regressor $\varphi(k)$ containing measured values, then the estimation error

$$[R(N)]^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \cdot \nu_0(k)$$

will also be small.

Important

If both the input $(u(k) \quad k = 1, 2, \ldots)$ and the error $(\nu_0(k) \quad k = 1, 2, \ldots)$ are stationary stochastic processes in an AR(MA)X model, then the output $(y(k) \quad k = 1, 2, \ldots)$ will also be a stationary process.

We want that the estimated model be close to the real model. Looking at equation of the estimated model $y(k) = p_0^\top \cdot \varphi(k) + \nu_0(k)$ we can see that it is close to the real model if the error $\nu_0(k)$ is as small as possible. Similarly the estimated parameter $\hat{p}_{LS}$ is close to the real parameter p_0 if the estimation error is small. The estimation error can be expressed as $[R(N)]^{-1} \frac{1}{N} \sum_{k=1}^{N} \varphi(k) \cdot \nu_0(k)$. We want to make this expression as small as possible. The estimation error $[R(N)]^{-1}$ contains the product of the regressor $\varphi(k)$ and the error $\nu_0(k)$. When the $\nu_0(k)$ error is small compared to the regressor $\varphi(k)$ containing measured values, then the estimation error will also be small.

stationary process: the statistical characteristics (mean, covariance,...) of the process do not change in the course of time t. The statistical characteristics at time t are the same as at time $t + \tau$ (similar to the time invariance of dynamic systems).

Conditions for the analysis of asymptotic behavior

(Assumptions for the analysis)

For the analysis of the asymptotic behaviour of the estimation error, let us assume that:

- *the error* $\{\nu_0(k)\}_{k=1}^N$ *is the realization of a stationary discrete time stochastic process*
- *the system itself can be described by an ARX model ,*
- *the input* $\{u(k)\}_{k=1}^N$ *is implemented as a stationary discrete time stochastic process .*

To be able to analyse the asymptotic behaviour of the estimation error, we need some assumptions:

- the error $\{\nu_0(k)\}_{k=1}^N$ is the realization of a stationary discrete time stochastic process
- the system itself can be described by an ARX model ,
- the input $\{u(k)\}_{k=1}^N$ is implemented as a stationary discrete time stochastic process .

From the first and third assumption it follows that the output is also a stationary discrete time stochastic process.

Entries of the $R(N)$ matrix

$$R(N) = \frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^{\top}(k)$$

where

$$\varphi(k) = [y(k-1)\ y(k-2)\ \ldots\ y(k-n)\ \ u(k)\ u(k-1)\ \ldots\ u(k-m)]^{\top}$$

(The elements of the $R(N)$ matrix)

$$R(N) = \frac{1}{N}\sum_{k=1}^{N}\left[\begin{array}{ccc|ccc} y(k-1)y(k-1) & \ldots & y(k-1)y(k-n) & y(k-1)u(k) & \ldots & y(k-1)u(k-m) \\ y(k-2)y(k-1) & \ldots & y(k-2)y(k-n) & y(k-2)u(k) & \ldots & y(k-2)u(k-m) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ y(k-n)y(k-1) & \ldots & y(k-n)y(k-n) & y(k-n)u(k) & \ldots & y(k-n)u(k-m) \\ - - - & - & - - - & - - - & - & - - - \\ u(k)y(k-1) & \ldots & u(k)y(k-n) & u(k)u(k) & \ldots & u(k)u(k-m) \\ u(k-1)y(k-1) & \ldots & u(k-1)y(k-n) & u(k-1)u(k) & \ldots & u(k-1)u(k-m) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ u(k-m)y(k-1) & \ldots & u(k-m)y(k-n) & u(k-m)u(k) & \ldots & u(k-m)u(k-m) \end{array}\right]$$

The elements of the $R(N)$ matrix can be seen here. From the product of $\varphi(k)$ and $\varphi^T(k)$ we get three different types of products:

- output x output e.g. $y(k-1)y(k-1)$
- input x output or output x input e.g. $y(k-1)u(k)$
- input x input $u(k)u(k)$

Asymptotic behaviour of the entries of $R(N) - 1$

The regressor $\varphi(\cdot)$ contains only earlier discrete time inputs and outputs in the AR(MA)X case, therefore the entries of $[R(N)]_{ij}$ can be divided into three classes

- *input autocovariance*: autocovariance function $r_{uu}(\tau)$ of the $\{u(k)\}_{k=1}^{N}$ stochastic process

$$\hat{R}_u^N(\tau) = \frac{1}{N}\sum_{k=1}^{N} u(k)\cdot u(k-\tau) \quad \rightarrow \quad R_u(\tau) = r_{uu}(\tau)$$

- *output autocovariance*: autocovariance function of the $r_{yy}(\tau)$ of the $\{y(k)\}_{k=1}^{N}$ stochastic process

$$\hat{R}_y^N(\tau) = \frac{1}{N}\sum_{k=1}^{N} y(k)\cdot y(k-\tau) \quad \rightarrow \quad R_y(\tau) = r_{yy}(\tau)$$

- *input-output covariance*: $r_{yu}(\tau)$ cross-covariance function of the two previous stochastic processes

$$\hat{R}_{yu}^N(\tau) = \frac{1}{N}\sum_{k=1}^{N} y(k)\cdot u(k-\tau) \quad \rightarrow \quad R_{yu}(\tau) = r_{yu}(\tau)$$

At time τ the elements of the $R(N)$ matrix go to a specific value when the number of samples goes to infinity. This follows from the assumption that the error, the input and the output are all stationary processes. Therefore their auto- and cross-covariances depend on the time difference (τ) only.

Asymptotic behavior of the entries of $R(N)$ – 2

Important

Consequently, the matrix $R(N)$ converges to a constant matrix R^* in the case of large sample size ($N \to \infty$).

(The error term)

The error process $\{\nu_0(k)\}_{k=1}^N$ is stationary, therefore

$$\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\nu_0(k) \quad \to \quad h^*$$

where h^* is a constant vector containing the elements of the cross-covariance functions of $\{u(k)\}_{k=1}^N$ and $\{\nu_0(k)\}_{k=1}^N$, or $\{y(k)\}_{k=1}^N$ and $\{\nu_0(k)\}_{k=1}^N$

Because all elements of $R(N)$ converges to a constant autocovariance or cross-covariance value, the matrix $R(N)$ converges to a constant matrix R^* in the case of large sample size ($N \rightarrow \infty$).
Because $\nu_0(k)$ is also a stationary process, therefore the $\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\nu_0(k)$ term also converges to a constant h^* vector, which contains the cross-covariances of $u(k)$ and $\nu_0(k)$, $y(k)$ and $\nu_0(k)$.

Conditions for asymptotic unbiasedness

Assume that the conditions for studying the asymptotic behavior are fulfilled. Then, the estimate is asymptotically unbiased, if:

- *matrix R^* is non-singular (*sufficient excitation*)*
  It is fulfilled if the processes $\{u(k)\}_{k=1}^N$ and $\{\nu_0(k)\}_{k=1}^N$ are independent and the R_{ij} composed of the $R_u(i-j)$ auto-correlations is non-singular (sufficiently exciting inputs).
- *$h^* = 0$ It is true if one of the below conditions are fulfilled:*
  - *The $\{\nu_0(k)\}_{k=1}^N$* error process is a white noise process with zero mean*: there is no modelling error, and the measurement error is white. Then the $\nu_0(k)$ error is independent of the past. Therefore, all the terms in $E\{\varphi(k)\cdot\nu_0(k)\}$ are zero.*
  - *The input $\{u(k)\}_{k=1}^N$ is a white noise process, and the order of the system $n = 0$: the actual output does not depend on past outputs. The regressor $\varphi(k)$ contains only the values of the past inputs, thus $E\{\varphi(k)\cdot\nu_0(k)\} = 0$.*

If the previous assumptions for the analysis of the asymptotic behaviour are fulfilled, then the estimate can be asymptotically unbiased in two cases. The estimation error converges to $[R^*]^{-1}h^*$, therefore

- R^* is non singular (otherwise the inverse does not exist). It is fulfilled if the processes $\{u(k)\}_{k=1}^N$ and $\{\nu_0(k)\}_{k=1}^N$ are independent and the R_{ij} composed of the $R_u(i-j)$ auto-correlations is non-singular (sufficiently exciting inputs)
- $h^* = 0$ It is true if one of the below conditions are fulfilled:
  - The $\{\nu_0(k)\}_{k=1}^N$ **error process is a white noise process with zero mean**: there is no modelling error, and the measurement error is white (no dynamics of the measurement system). Then the $\nu_0(k)$ error is independent of the past. =¿the noise is not correlated with $\varphi(k)$. Therefore, all the terms in $E\{\varphi(k)\nu_0(k)\}$ are zero ($E(XY) = E(X)E(Y)$ if X and Y are independent).
  - The input $\{u(k)\}_{k=1}^N$ is a white noise process, and *the order of the system* $n = 0$: the actual output does not depend on past outputs. The regressor $\varphi(k)$ contains only the values of the past inputs: $(\varphi(k) = [u(k), u(k-1), \ldots, u(k-d-m)])$, thus $E\{\varphi(k) \cdot \nu_0(k)\} = 0$ ($u(k)$ and $\nu_0(k)$ are independent). If the order of the model is not zero, then the regressor contains past output data (e.g. y(k-1)), which contains $\nu_0(k-1)$ that is correlated with $\nu_0(k)$

Properties of LS estimates of predictive models

Important (Asymptotic distribution of the estimate)

If the Conditions for asymptotic unbiasedness are fulfilled, then the distribution of the random variable

$$\sqrt{N} \cdot (\hat{p}_{LS}(N) - p_0)$$

will be a multi-dimensional Gaussian distribution *with* 0 *mean.*

(Covariance matrix of the estimate)

The covariance matrix in the SISO case is $\lambda_0 [R^*]^{-1}$*, where* λ_0 *is the variance of the* $\{\nu_0(k)\}_{k=1}^N$ *error.*

SUMMARY: LS estimate of ARX model parameters

Important (Steps of the estimation)

1. *Collect the data* $D^N = \{(y(k), u(k)), k = 1, .., N\}$ *using a white noise input sequence and form the regressor vectors for* $k = 1, ..., N$

$$\varphi(k) = [y(k-1)\ y(k-2)\ \ldots\ y(k-n)\ \ u(k)\ u(k-1)\ \ldots\ u(k-m)]^\top$$

2. *LS estimate of the parameters* $p = [-a_1\ -a_2\ \ldots\ -a_n\ \ b_0\ b_1\ \ldots\ b_m]^\top$*:*

$$\hat{p}_{LS} = \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\varphi^\top(k)\right]^{-1}\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot y(k)$$

3. *From the prediction error sequence* $\varepsilon(k) = y(k) - \hat{y}(k|p) = y(k) - \hat{p}_{LS}^\top\varphi(k)$ *calculate its estimated variance* λ_0
4. *Covariance matrix of the estimated parameters* $\hat{p}_{LS}$

$$C\hat{O}V\{\hat{p}_{LS}\} = \lambda_0 \cdot \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\cdot\varphi^\top(k)\right]^{-1}$$

The steps of the estimation of a dynamic ARX model parameters are here:

1. The white noise input is required for the sufficient excitation. Without sufficient excitation the asymptotic unbiasedness of the estimate is not guaranteed.
2. The parameters can be estimated with the LS method.
3. To evaluate the quality of the estimate, compute the prediction error sequence (residuals), which is the difference between the measured output ($y(k)$) and the estimated model output ($\hat{p}_{LS}^{\top}\varphi(k)$). We can also calculate the variance of the prediction error sequence, which will be used to compute the covariance matrix of the estimated parameters. The prediction error sequence should be a white noise.
4. The covariance matrix of the estimate can be computed using the variance of the prediction error sequence and the regressor. The non-diagonal elements of the covariance matrix indicate the dependency the parameters. They should be as small as possible.

Overview

Overview of data

With a careful overview of data, we can recognize the following phenomena:

- trends
- outliers
- apparent errors

(Visual overview)

For the visual overview, we should plot the data:

- *as a function of time (as data sequences)*
- *as a function of each other*

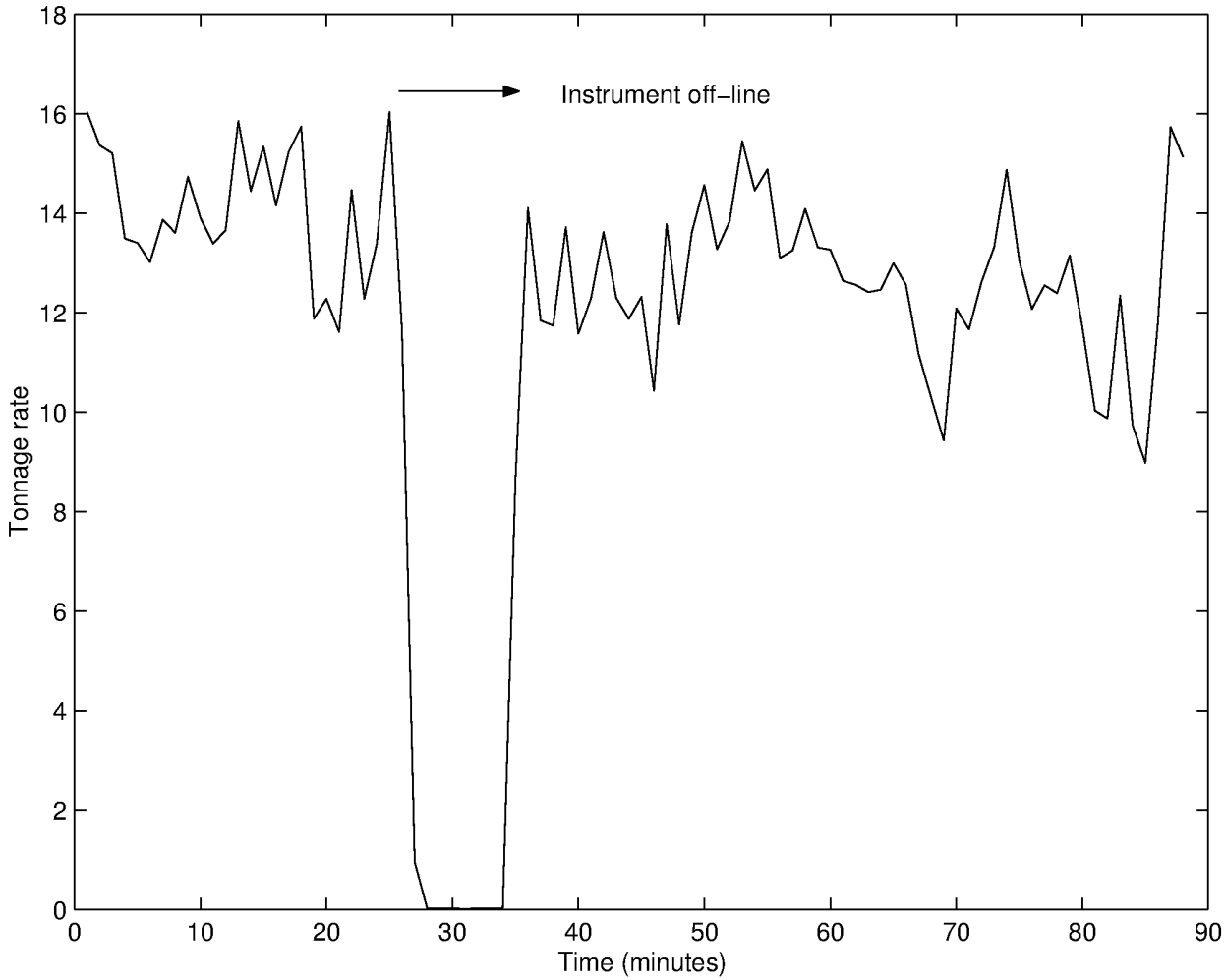
The quality of the data affects the quality of the estimation. Using 'bad' data, it may occur that the estimation cannot be carried out, or the result will be not good (e.g. biased estimate). Therefore it is important to examine the data before starting the estimation process. The aim of the overview of the data is to determine whether it is suitable for parameter estimation or not. It is recommended to discard the data, which is not suitable and repeat the measurement under better conditions. Trying to correct the bad data is usually time-consuming and it is not sure that we correct the data well.The following phenomena should be recognized:

- trends: slow, long term variance in the data. different types of trends: linear, polynomial, exponential, etc.
- outliers: measured data is significantly larger or smaller than it expected
- apparent error: e.g. missing data, sensor error, abrupt changes

The method of data overview should be as simple as possible. Usually the visual overview of the data is enough to decide its quality. The measured data is often plotted in graphs as a function of time or as a function of each other.

Visual overview - Serious error

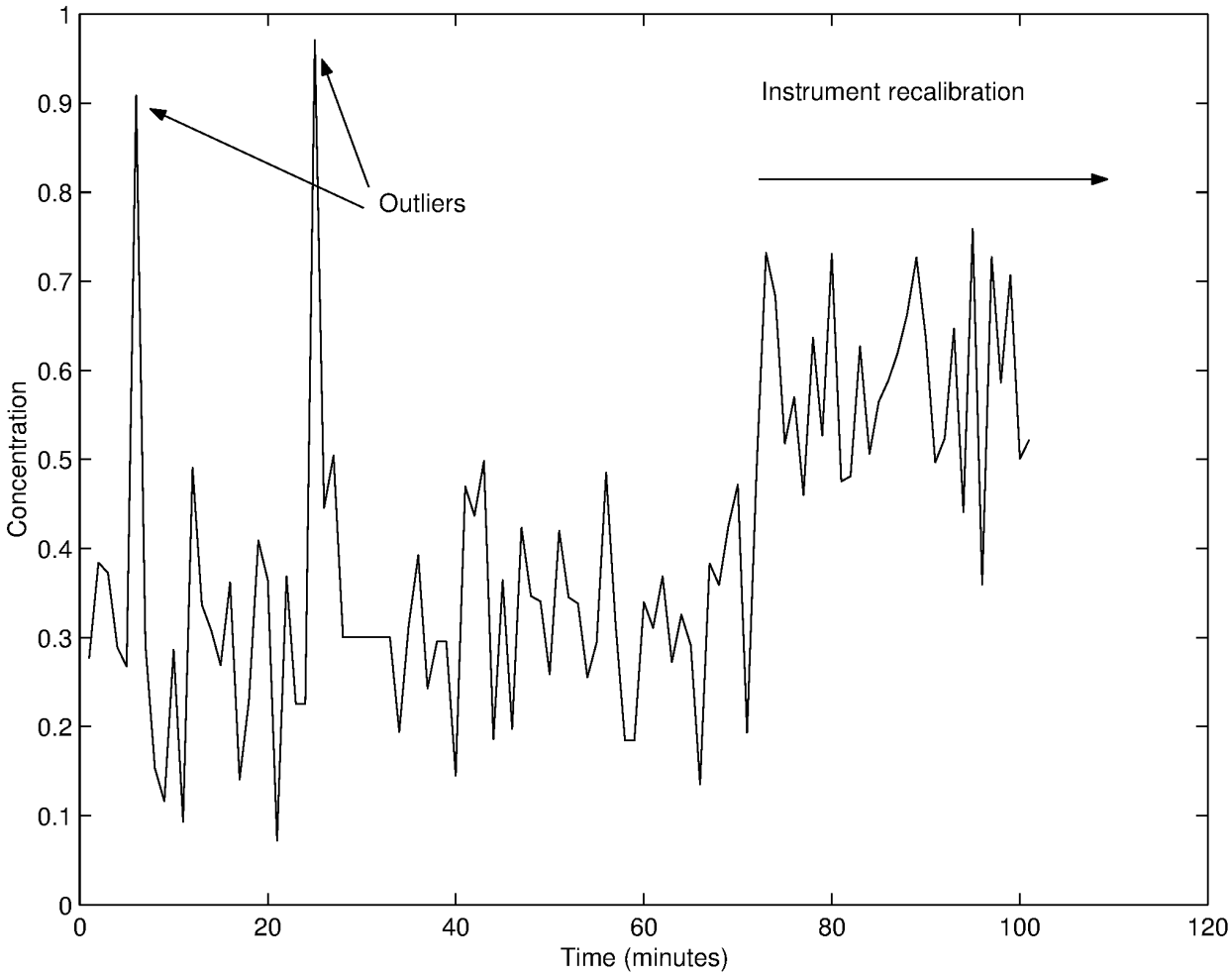
Measurement with serious error:



In this figure an example of serious error is shown. The measurement was interrupted around 30 minutes, because of the instrument/sensor break. There is no measurement data at this time interval.

Visual overview - Outliers and gross error

Measurement with outliers and error:



In this figure examples of outliers and a gross error can be seen. The outlier data are much greater than the other measured data. From time 70, a change in the mean of the measured data can be observed, which was caused by the recalibration of the sensor.

Possible causes of trends

Important (Monitoring of trends and steady states)

Most frequent causes of trends in measured data:

- *fault of sensors that might be indicated by a slow drift-like change*
- *slow, unmodeled process (ageing, equipment deterioration, equipment becomes dirty) indicated by a drift-like change, too*
- *slow, usually periodic disturbance: seasonal, weekly or daily variation (e.g. temperature), effects of change of shifts, weekend, different operation of night-shifts etc.*

The simplest trend detection method is fitting a line to the data. If the slope of the line is not zero, then there is a trend in the data, which can be easily removed. Non-linear trends can be removed, too.

Outliers

Important

Data generated by Gaussian random variables may (theoretically) contain arbitrarily large or small values with nonzero (but very small) probability.

(Notion of outliers)

In practice, a measured data point can be considered an outlier if its relative magnitude, i.e. $||d(i) - \overline{d}||$ *is significantly larger than the deviation of the measurement errors, where* $\overline{d}$ *is the mean and* $||.||$ *is a suitable vector norm.*

How to determine*:*

- *by simple visual overview of the data*
- *check normality of data using e.g.* χ^2 *test*

What is an outlier? Theoretically data generated by Gaussian random variables may contain arbitrarily large or small data, with very small probability. Therefore a very large or very small data may be the result of the normal operation. In practice we consider a measured data point an outlier, if its deviation from the sample mean is much greater than the sample deviation. The deviation from the sample mean is measured by a suitable norm.

The outliers can be usually determined by simple visual overview.

More difficult mathematical methods are also exist (χ^2 test, assuming normally distributed noise).

Overview

Experiment design

Aim: to determine the optimal input for parameter estimation

- asymptotic unbiasedness
- minimal variance, uncorrelated elements

Important (Experiment parameters to be chosen)

- *sampling time*
- *number of samples*
- *test signals for sufficient excitation*

The measured data can be obtained from 'passive' measurements, when we cannot intervene to the process (e.g. an atomic reactor). However, in most cases we can create special, artificial input sequences and collect the measured outputs of the system. The aim of the experiment design is to determine the optimal input for the parameter estimation such that the estimated parameters have the best statistical properties.
The estimated parameter(s) should be asymptotically unbiased, it has minimal variance, and the parameters are uncorrelated.

Experiment design usually covers the selection of the input signal (to provide sufficient excitation), the sampling time and number of samples (-¿ how long the measurement lasts).

Choosing the sampling time

(Aims)

We should aim at

- *provide sufficiently high frequency sampling for sufficiently long time,*
- *the sample should contain enough information for each important and modelled time constant (pole) of the system*

- *sampling time should be about 1/4 of (or smaller than) the fastest (smallest) time constant*
- *measurement time should be at least 4 times larger than the slowest time constant*

Most dynamic systems are continuous time systems, but the parameter estimation is executed using discrete time systems. Therefore the continuous system need to be sampled with a specific sample time and duration. When choosing the sampling time we need to take care of the following:

- frequency of the sampling: fast changes in the signals can be caught with high frequency sampling. However high frequency sampling of a slowly changing process generates unnecessarily much data (large size of the measurement file). The duration of the measurement should be long enough to record all important changes in the signals
- All types of dynamics of the system is need to be caught. The measured data should contain enough information about the smallest and the largest time constants too.

Choosing the the number of samples

The number of measurements needed for parameter estimation depends:

- *on the number of measurements in one record*
- *on the number of repetitions (records)*

Important

The overall number of samples should be significantly larger than the order of the system and also much larger than the number of estimated parameters.

If we repeat the measurements with the same input signal, then we will get more information about the measurement noise, which is important for the quality evaluation of estimates.

Sufficient excitation

Test signals for sufficient excitation

Main considerations:

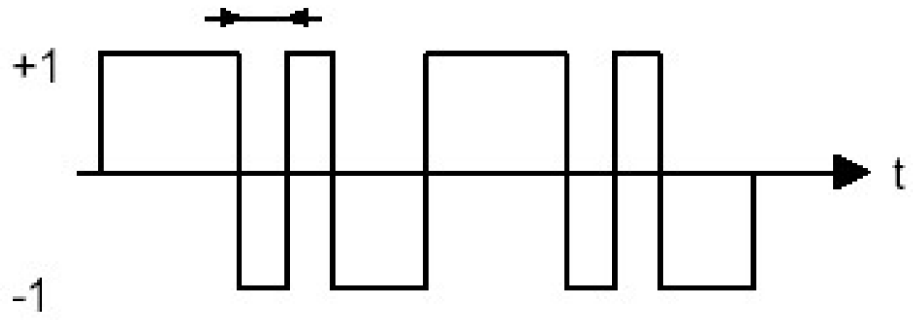
- *Appropriate signal to noise ratio*
  For this, a suitably chosen test-signal is often added to the normal input of the system to ensure sufficient excitation
- *Asymptotic unbiasedness*
  The inputs should be independent from the other noises and disturbances. Moreover, it is advantageous if the input is (approximately) white noise.

The system dynamics should be 'sufficiently' excited, Noises/disturbances are usually present in all systems, but they are too small to affect the system behaviour. To ensure the sufficient excitation of the system, we use special test signals.

At the conditions of asymptotic unbiasedness, it was required that the input be independent from the other noises and disturbances, and be a white noise signal in the **dynamic** case. In practice it is enough if the input is approximately white noise.

The PRBS test signal

It has only two values, and jumps randomly between them
(takes the value $+1$ with the probability p)



white noise process with binomial distribution

Important (Sampling time and number of samples)

sampling time should be 1/4 - 1/5 of the smallest time constant
number of samples should be 4-5 times the largest time constant

$$x(t) = \begin{cases} x_{min}, & \text{if} \mu < 0.5 \\ x_{max}, & \text{if} \mu \geq 0.5 \end{cases}$$

where μ is a uniformly distributed random variable on the interval (0,1). The Pseudo Random Binary Sequence is a commonly used test signal, that has two values and changes arbitrarily between them. It provides sufficient excitation, but requires relatively long measurement time (lots of measurement points) and it does not disturb the normal system operation a lot. The design parameters of the PRBS are the

- the two values, which the signal can have
- the sampling time (i.e.clock period/frequency of the signal)
- the number of samples, i.e. the length of the PRBS signal

Overview

Quality of the estimate – 1

Analysis of the residuals

Residual: realization of the prediction error series

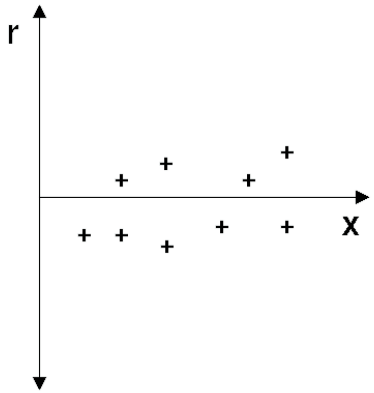
$$\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta) \quad , \quad k = 1, \ldots, N$$

Important

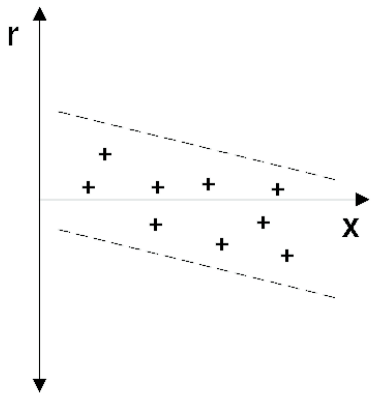
For an unbiased estimation, the residuals are uncorrelated and have 0 mean.

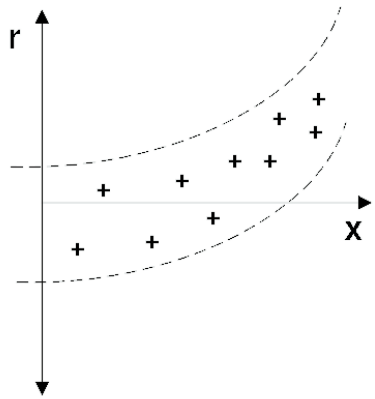
the prediction error sequence (residuals) gives information about how well the estimated model fits the measured data. If the residual sequence is uncorrelated and has 0 mean then the difference between the measured and the estimated data is caused by only the measurement noise.
If the mean is not zero, but the prediction error is white noise, then a constant term is probably missing from the model.
If the prediction error sequence is not white, then the goodness of the model depends on the time, meaning that the measurements and the noise are possbly correlated.

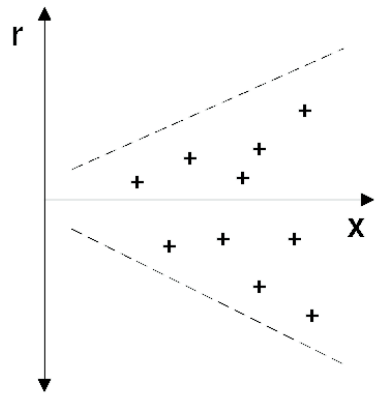
Residuals



random







non random

Possible residuals

The residuals in the top figure are good.
The bottom left residuals have a linear trend.
The bottom middle residuals have nonlinear trend.
The variance of the residuals in the bottom right figure is constantly growing, i.e. the model becomes less and less reliable as the time increases.

Testing the zero mean property

Detecting trends:

- **fitting a linear function to the data**
- standard statistical analysis for data distribution (in the case of independent, identically distributed measurement errors)
- cumulative sum (CUSUM) method (recursive mean):

$$s[k] = \frac{1}{k}\sum_{i=1}^{k} d(i) = \frac{1}{k}\left((k-1)s[k-1] + d(k)\right)$$

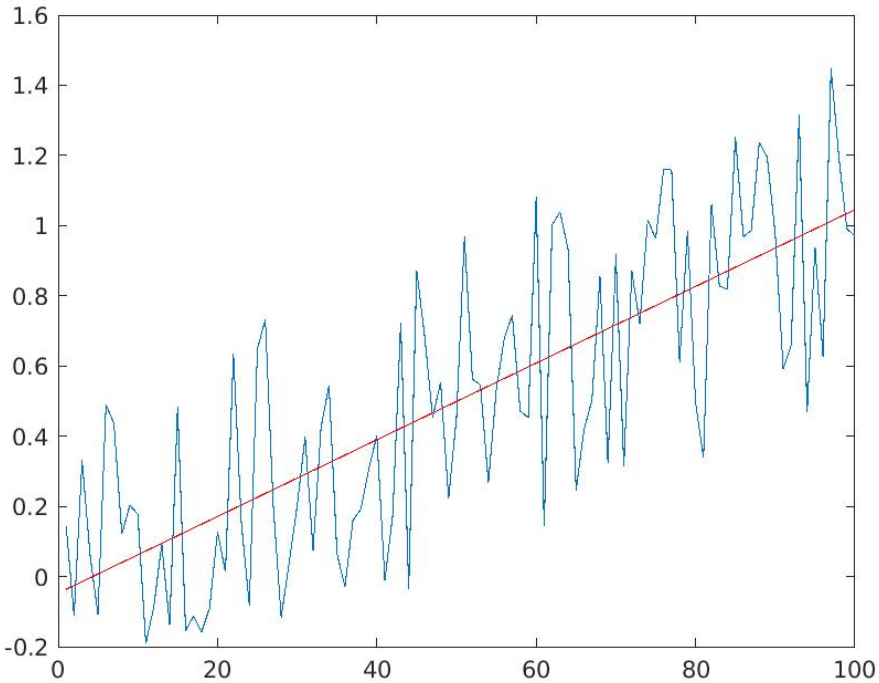
The computed $s[k]$ is plotted as a function of time (k), and the trend is monitored. The variance of $s[k]$ decreases with the increase of measurement data.

The zero mean property of the residuals can be checked by

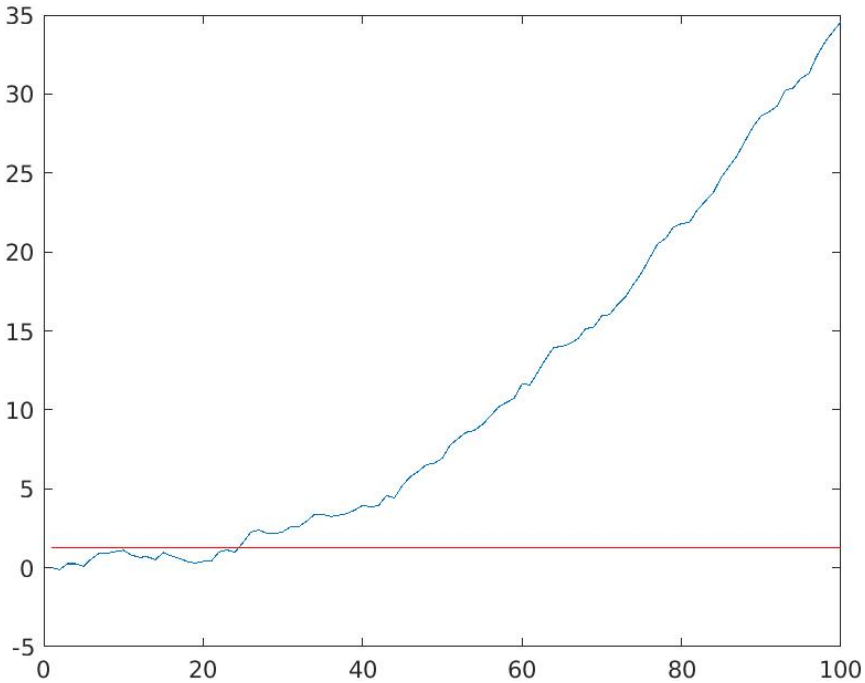
- Fitting a line to the data. If there is no trend in the data, then the slope of the line is 0.
- CUSUM (cumulative sum) is used for detecting changes in the mean. s[k] is the sample mean up to k. When the cumulative mean exceeds a certain treshold (e.g. 5*standard deviation of the sample) then a change is detected.

Testing the zero mean property

Detecting trends



fitting a linear function



cumulative sum

Recall Estimating the mean value and the variance

Assume that the underlying random variable ξ has a mean value m and the variance σ^2

- Mean value

  statistics is the *sample mean*

  $$\mu(S) = \frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n) \quad , \quad \hat{m}(D) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

  Property: $E[\mu] = m$
- Variance

  statistics is the *corrected empirical variance*

  $$\theta(S) = \frac{1}{n-1}\left((\xi_1 - \mu)^2 + (\xi_2 - \mu)^2 + ... + (\xi_n - \mu)^2\right)$$

  Property: $E[\theta] = \sigma^2$
- Unbiased estimate

  if the *mean value of the statistics is the real value of the parameter to be estimated*

Recall Estimation of the covariances

Consider (scalar valued) random variables ξ_i *from the same distribution but* not independent. *They form a "generalized" sample* $S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$.

Estimation of the mean value m

- Estimate

$$\hat{m}(D) = \frac{1}{n}(x_1 + ... + x_n)$$

- It may be a biased estimate

Estimation of the auto-covariances $r_{\xi\xi}(s),\ s = 0, 1, ...$

- Estimate for $s << n$

$$\hat{r}(D) = \frac{1}{n-s}\left((x_1 - \hat{m})(x_{s+1} - \hat{m}) + ... + (x_{n-s} - \hat{m})(x_n - \hat{m})\right)$$

- It may be a biased estimate

Important

Variance and correlation: *only for data samples without trends!*

Consider scalar valued measured data $D[1,k] = D^k$ *of* k *measurements:* $d_1, ..., d_k$. *They form a "generalized" sample*

$$S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\} \quad \sim \quad S(D^k) = \{d_1, d_2, ..., d_n\}$$

Important

Variance and correlation computation: *from the generalized sample* $S(D^k)$.

Testing the independence

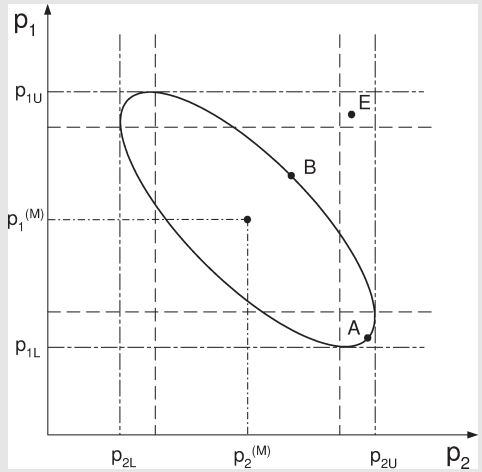
- Residuals vs. other variables
  - Plot the residuals against any time variables present or used (e.g. regressors). A non-random pattern implies dependence.
- Residuals with each other
  - Compute and plot the autocorrelation sequence. It shows the dependence between the subsequent residuals. Random pattern is expected.

Quality of the estimate in the parameter space

Analysis of the covariance matrix*: estimate*

$$\widehat{COV}(\hat{p}_{LS}) = \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^T(k)\right]^{-1}\Delta_\varepsilon$$

For a 'good' estimate, the parameter values are uncorrelated



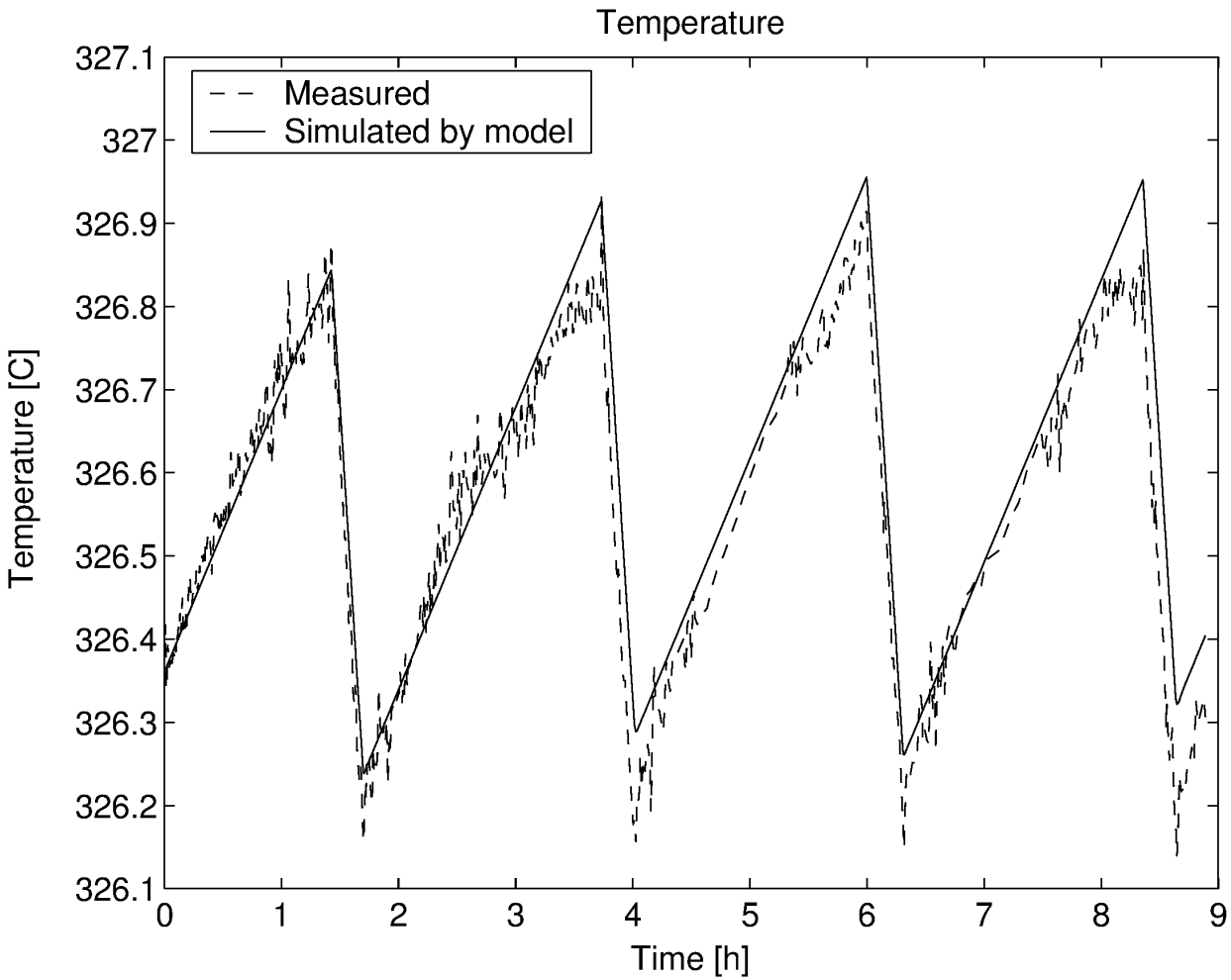
Confidence intervals

The covariance matrix of $\hat{p}_{LS}$ can be determined by the properties of the regressors ($\varphi(k)$) and the variance of the residuals Δ_ε. When we have the chance to manipulate $\varphi(k)$, it is an important experiment design issue to make the inverse small .Δ_ε is ususally not known therefore it is important to estimate it from data.

The confidence intervals represents the interval where the parameter values falls with a certain (usually 95%) probability.

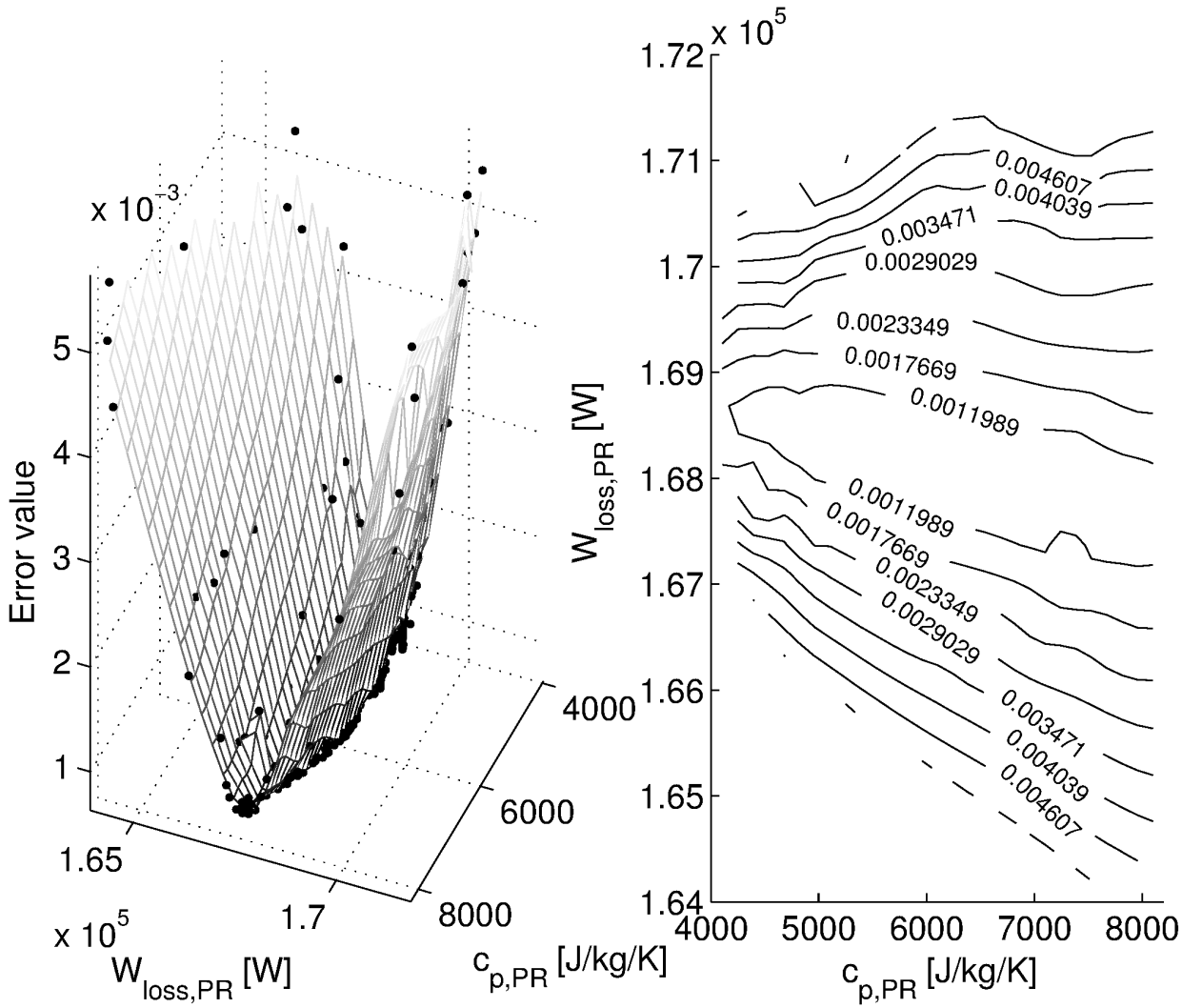
In nonlinear cases the confidence intervals can be estimated by the level sets of the loss function.

Example: quality of the estimate, prediction error



Measured and model computed (predicted) data

Example: estimated confidence intervals



Level sets of the loss function