Introduction

The temporal behavior of systems, such as e.g. technical systems from the areas of electrical engineering, mechanical engineering, and process engineering, as well as non-technical systems from areas as diverse as biology, medicine, chemistry, physics, economics, to name a few, can uniformly be described by mathematical models. This is covered by systems theory. However, the application of systems theory requires that the mathematical models for the static and dynamic behavior of the systems and their elements are known. The process of setting up a suitable model is called modeling. As is shown in the following section, two general approaches to modeling exist, namely theoretical and experimental modeling, both of which have their distinct advantages and disadvantages.

1.1 Theoretical and Experimental Modeling

A system is understood as a confined arrangement of mutually affected entities, see e.g. DIN 66201. In the following, these entities are processes. A process is defined as the conversion and/or the transport of material, energy, and/or information. Here, one typically differentiates between individual (sub-)processes and the entire process. Individual processes, i.e. (sub-)processes, can be the generation of mechanical energy from electric energy, the metal-cutting machining of workpieces, heat transfer through a wall, or a chemical reaction. Together with other sub-processes, the entire process is formed. Such aggregate processes can be an electrical generator, a machine tool, a heat exchanger, or a chemical reactor. If such a process is understood as an entity (as mentioned above), then multiple processes form a system such as e.g. a power plant, a factory, a heating system, or a plastic material production plant. The behavior of a system is hence defined by the behavior of its processes.

The derivation of mathematical system and process models and the representation of their temporal behavior based on measured signals is termed system analysis respectively process analysis. Accordingly, one can speak of system identification or process identification when applying the experimental system or process analysis techniques described in this book. If the system is excited by a stochastic signal,
one also has to analyze the signal itself. Thus the topic of signal analysis will also be treated. The title Identification of Dynamic Systems or simply Identification shall thus embrace all areas of identification as listed above.

For the derivation of mathematical models of dynamic systems, one typically discriminates between theoretical and experimental modeling. In the following, the basic approach of the two different ways of modeling shall be described shortly. Here, one has to distinguish lumped parameter systems and distributed parameter systems.
The states of *distributed parameter systems* depend on both the time and the location and thus their behavior has to be described by *partial differential equations (PDEs)*. *Lumped parameter systems* are easier to examine since one can treat all storages and states as being concentrated in single points and not spatially distributed. In this case, one will obtain *ordinary differential equations (ODEs)*.

For the *theoretical analysis*, also termed *theoretical modeling*, the model is obtained by applying methods from calculus to equations as e.g. derived from physics. One typically has to apply simplifying assumptions concerning the system and/or process, as only this will make the mathematical treatment feasible in most cases. In general, the following types of equations are combined to build the model, see also Fig. 1.1 (Isermann, 2005):

1. *Balance equations*: Balance of mass, energy, momentum. For distributed parameter systems, one typically considers infinitesimally small elements, for lumped parameter systems, a larger (confined) element is considered
2. *Physical or chemical equations of state*: These are the so-called constitutive equations and describe reversible events, such as e.g. inductance or the second Newtonian postulate
3. *Phenomenological equations*: Describing irreversible events, such as friction and heat transfer. An entropy balance can be set up if multiple irreversible processes are present
4. *Interconnection equations* according to e.g. Kirchhoff’s node and mesh equations, torque balance, etc.

By applying these equations, one obtains a set of ordinary or partial differential equations, which finally leads to a theoretical model with a certain structure and defined parameters if all equations can be solved explicitly. In many cases, the model is too complex or too complicated, so that it needs to be simplified to be suitable for subsequent application. Figure 1.2 shows the order of the execution of individual simplifying actions. The first steps of this simplification procedure can already be carried out as the fundamental equations are set up by making appropriate simplifying assumptions. It is very tempting to include as many physical effects into the model as possible, especially nowadays, where simulation programs offer a wide variety of pre-build libraries of arbitrary degrees of complexity. However, this often occludes the predominant physical effects and makes both the understanding and the work with such a model a very tiresome, if not infeasible, endeavor.

But even if the resulting set of equations cannot be solved explicitly, still the individual equations give important hints concerning the model structure. Balance equations are always linear, some phenomenological equations are linear in a wide range. The physical and chemical equations of state often introduce non-linearities into the system model.

In case of an *experimental analysis*, which is also termed *identification*, a mathematical model is derived from measurements. Here, one typically has to rely on certain a priori assumptions, which can either stem from theoretical analysis or from previous (initial) experiments, see Fig. 1.1. Measurements are carried out and the input as well as the output signals are subjected to some identification method in order
to find a mathematical model that describes the relation between the input and the output. The input signals can either be a section of the natural signals that act on the process during normal operation or can be an artificially introduced test signal with certain prespecified properties. Depending on the application, one can use parametric or non-parametric models, see Sect. 1.2. The resulting model is termed experimental model.

The theoretically and the experimentally derived models can be compared if both approaches can be applied and have been pursued. If the two models do not match, then one can get hints from the character and the size of the deviation, which steps of the theoretical or the experimental modeling have to be corrected, see Fig. 1.1.

Theoretical and experimental models thus complement one another. The analysis of the two models introduces a first feedback loop into the course of action for system analysis. Therefore, system analysis is typically an iterative procedure. If one is not interested in obtaining both models simultaneously, one has the choice between the experimental model (case A in Fig. 1.1) and the theoretical model (case B in Fig. 1.1). The choice mainly depends on the purpose of the derived model:

The theoretical model contains the functional dependencies between the physical properties of a system and its parameters. Thus, this model will typically be preferred if the system shall already be optimized in its static and dynamic behavior during the

Fig. 1.2. Basic approach for theoretical modeling
design phase or if its temporal behavior shall be simulated prior to the construction, respectively completion of the system.

On the contrary, the experimental model does only contain numbers as parameters, whose functional relations to the process properties remain unknown. However, this model can describe the actual dynamics of the system better and can be derived with less effort. One favors these experimental models for the adaptation of controllers (Isermann, 1991; Isermann et al, 1992; Åström et al, 1995; Åström and Wittenmark, 1997) and for the forecast of the respective signals or fault detection (Isermann, 2006).

In case B (Fig. 1.1), the main focus is on the theoretical analysis. In this setting, one employs the experimental modeling only once to validate the fidelity of the theoretical model or to determine process parameters, which can otherwise not be determined with the required accuracy. This is noted with the sequence B/1 in Fig. 1.1.

In contrast to case B, the emphasis is on the experimental analysis in case A. Here, one tries to apply as much a priori knowledge as possible from the theoretical analysis, as the model fidelity of the experimental model normally increases with the amount of a priori knowledge exploited. In the ideal case, the model structure is already known from the theoretical analysis (path A/2 in Fig. 1.1). If the fundamental equations of the model cannot be solved explicitly, if they are too complicated, or if they are not even completely known, one can still try to obtain information about the model structure from this incomplete knowledge about the process (sequence A/1 in Fig. 1.1).

The preceding paragraphs already pointed out that the system analysis can typically neither be completely theoretical nor completely experimental. To benefit from the advantages of both approaches, one does rarely use only theoretical modeling (leading to so-called white-box models) or only experimental modeling (leading to so-called black-box models), but rather a mixture of both leading to what is called gray-box models, see Fig. 1.3. This is a rather suitable combination of the two approaches, which is determined by the scope of application of the model and the system itself. The scope of application defines the required model accuracy and hence the effort that has to be put into the analysis. This introduces a second feedback loop into the schematic diagram presented in Fig. 1.1, which starts at the resulting models (either theoretical or experimental) and goes back to the individual modeling steps, hence one is confronted with a second iteration loop.

Despite the fact that the theoretical analysis can in principle deliver more information about the system, provided that the internal behavior is known and can be described mathematically, experimental analysis has found ever increasing attention over the past 50 years. The main reasons are the following:

- Theoretical analysis can become quite complex even for simple systems
- Mostly, model coefficients derived from the theoretical considerations are not precise enough
- Not all actions taking place inside the system are known
The actions taking place cannot be described mathematically with the required accuracy.
Some systems are very complex, making the theoretical analysis too time-consuming.
Identified models can be obtained in shorter time with less effort compared to theoretical modeling.

The experimental analysis allows the development of mathematical models by measurement of the input and output of systems of arbitrary composition. One major advantage is the fact that the same experimental analysis methods can be applied to diverse and arbitrarily complex systems. By measuring the input and output only, one does however only obtain models governing the input-output behavior of the system, i.e. the models will in general not describe the precise internal structure of the system. These input-output models are approximations and are still sufficient for many areas of application. If the system also allows the measurement of internal states, one can obviously also gather information about the internal structure of the
Table 1.1. Properties of theoretical modeling and identification

<table>
<thead>
<tr>
<th>Theoretical Modeling</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model structure follows from laws of nature</td>
<td>Model structure must be assumed</td>
</tr>
<tr>
<td>Modeling of the input/output behavior as well as the internal behavior</td>
<td>Only the input/output behavior is identified</td>
</tr>
<tr>
<td>Model parameters are given as function of system properties</td>
<td>Model parameters are “numbers” only, in general no functional dependency to system properties known</td>
</tr>
<tr>
<td>Model is valid for the entire class of processes of a certain type and for different operating conditions</td>
<td>Model is only valid for investigated system and within operating limits</td>
</tr>
<tr>
<td>Model coefficients are not known exactly</td>
<td>Model coefficients are more precise for the given system within operating limits</td>
</tr>
<tr>
<td>Models can be build for non-existing systems</td>
<td>Model can only be identified for an existing system</td>
</tr>
<tr>
<td>The internal behavior of the system must be known and must be describable mathematically</td>
<td>Identification methods are independent of the investigated system and can thus be applied to many different systems</td>
</tr>
<tr>
<td>Typically lengthy process which takes up much time</td>
<td>Fast process if identification methods exist already</td>
</tr>
<tr>
<td>Models may be rather complex and detailed</td>
<td>Model size can be adjusted according to the area of application of the model</td>
</tr>
</tbody>
</table>

system. With the advent of digital computers starting in the 1960s, the development of capable identification methods has started. The different properties of theoretical modeling and identification have been summarized and set in contrast in Table 1.1.

1.2 Tasks and Problems for the Identification of Dynamic Systems

A process with a single input and a single output (SISO) is considered in the following. The process shall be stable to ensure a unique relation between input and output. Both the input and the output shall be measured without error. The task of identifying the process $P$ is to find a mathematical model for the temporal behavior of the process from the measured input $u(t) = u_M(t)$, the measured output $y(t) = y_M(t)$ and optionally additional measured signals, see Fig. 1.4. This task is made more complicated, if disturbances $z_1 \ldots z_i$ are acting on the process and are influencing the output signal. These disturbances can have various causes. The disturbances seen in the measured signals often stem from noise and hence will also be included in the term noise in the remainder of this book. The output is thus corrupted by a noise $n(t)$. In this case, one has to apply suitable techniques to separate the wanted signal
$y_u(t)$, i.e. the response of the system due to the input $u(t)$, from the disturbances $n(t)$.

The term *identification* and the required subsequent tasks can thus be defined as follows:

Identification is the experimental determination of the temporal behavior of a process or system. One uses measured signals and determines the temporal behavior within a class of mathematical models. The error (respectively deviation) between the real process or system and its mathematical model shall be as small as possible.

This definition stems from Zadeh (1962), see also (Eykhoff, 1994). The measured signals are typically only the input to the system and the output from the system. However, if it is also possible to measure states of the process, then one can also gather information about the internal structure of the process.

In the following, a linear process is considered. In this case, the individual disturbance components $z_1, \ldots, z_i$ can be combined into one representative disturbance $n(t)$, which is added to the wanted signal $y_u(t)$, see Fig. 1.5. If this disturbance $n(t)$ is not negligibly small, then its counterfeiting influence must be eliminated by the identification method as much as possible. For decreasing signal-to-noise ratios, the measurement time $T_M$ must typically be increased.

For the identification itself, the following limitations have to be taken into consideration:

1. The available measurement time $T_M$ is always limited, either due to technical reasons, due to time variance of the process parameters or due to economical reasons (i.e. budget), thus

$$T_M \leq T_{M,max} \quad (1.2.1)$$

2. The maximum allowable change of the input signal, i.e. the test signal height $u_0$ is always limited, either due to technical reasons or due to the assumption of linear process behavior which is only valid within a certain operating regime

$$u_{\min} \leq u(t) \leq u_{\max} \quad (1.2.2)$$

3. The maximum allowable change of the output signal, $y_0$, may also be limited due to technical reasons or due to the assumption of linear process behavior
1.2 Tasks and Problems for the Identification of Dynamic Systems

Fig. 1.6. Examples of disturbance components. (a) high frequent quasi-stationary stochastic disturbance. (b) low-frequent non-stationary stochastic disturbance. (c) disturbance with unknown character

which again is only valid within a certain operating regime

\[ y_{\text{min}} \leq y(t) \leq y_{\text{max}} \]  

(1.2.3)

4. The disturbance \( n(t) \) typically consists of different components, which can be classified according to the following groups, see also Fig. 1.6:

a) High-frequent quasi-stationary stochastic noise \( n(t) \) with \( E\{n(t)\} = 0 \). Higher frequent deterministic signal with \( n(t) = 0 \).

b) Low-frequent non-stationary stochastic or deterministic signal (e.g. drift, periodic signals with period times of one day or one year) \( d(t) \)

c) Disturbance signal of unknown character (e.g. outliers) \( h(t) \)

It is assumed that within the limited measurement time, the disturbance component \( n(t) \) can be treated as a stationary signal. The low-frequent component \( d(t) \) must be treated as non-stationary, if it has stochastic character. Low-frequent deterministic disturbances can be drift and periodic signals with long period times such as one day or one year. Disturbance components with unknown character \( h(t) \) are random signals, which cannot be described as stationary stochastic signals even for long measurement periods. This can be e.g. suddenly appearing, persistent, or disap-
pearing disturbances and so-called outliers. These disturbances can e.g. stem from electromagnetic induction or malfunctions of the measurement equipment.

Typical identification methods can only eliminate the noise \( n(t) \) as the measurement time is prolonged. Simple averaging or regression methods are often sufficient in this application. The components \( d(t) \) require more specifically tailored measures such as special filters or regression methods which have been adapted to the very particular type of disturbance. Almost no general hints can be given concerning the elimination of the influence of \( h(t) \). Such disturbances can only be eliminated manually or by special filters.

Effective identification methods must thus be able to determine the temporal behavior as precisely as possible under the constraints imposed by

- the given disturbance \( y_z(t) = n(t) + d(t) + h(t) \)
- the limited measurement time \( T_M \leq T_{M,\text{max}} \)
- the confined test signal amplitude \( u_{\text{min}} \leq u(t) \leq u_{\text{max}} \)
- the constrained output signal amplitude \( y_{\text{min}} \leq y(t) \leq y_{\text{max}} \)
- the purpose of the identification.

Figure 1.7 shows a general sequence of an identification. The following steps have to be taken:

First, the purpose has to be defined as the purpose determines the type of model, the required accuracy, the suitable identification methods and such. This decision is typically also influenced by the available budget, either the allocated financial resources or the expendable time.

Then, a priori knowledge must be collected, which encompasses all readily available information about the process to be identified, such as e.g.

- recently observed behavior of the process
- physical laws governing the process behavior
- rough models from previous experiments
- hints concerning linear/non-linear, time-variant/time-invariant as well as proportional/integral behavior of the process
- settling time
- dead time
- amplitude and frequency spectrum of noise
- operating conditions for conduction of measurements.

Now, the measurement can be planned depending on the purpose and the available a priori knowledge. One has to select and define the

- input signals (normal operating signals or artificial test signals and their shape, amplitude and frequency spectrum)
- sampling time
- measurement time
- measurements in closed-loop or open-loop operation of the process
- online or offline identification
- real-time or not
Task Budget

Purpose

Planning of Measurements

Signal Generation Measurement Results

Data Inspection and Preprocessing

Application of Identification Method

Process Model Non-Parametric Parametric

Model Validation

Yes

A-Priori Knowledge

Assumption of Model Structure

Determination of Model Structure

No

Resulting Model

Fig. 1.7. Basic sequence of the identification
Introduction

Once these points have been clarified, the measurements can be conducted. This includes the signal generation, measurement, and data storage.

The collected data should undergo a first visual inspection and outliers as well as other easily detectable measurement errors should be removed. Then, as part of the further pre-processing, derivatives should be calculated, signals be calibrated, high-frequent noise be eliminated by e.g. low-pass filtering, and drift be removed. Some aspects of disturbance rejection and the removal of outliers by graphical and analytical methods are presented in Chap. 23. Methods to calculate the derivatives from noisy measurements are shown in Chap. 15.

After that, the measurements will be evaluated by the application of identification techniques and determination of model structure.

A very important step is the performance evaluation of the identified model, the so-called validation by comparison of model output and plant output or comparison of the experimentally established with the theoretically derived model. Validation methods are covered in Chap. 23. Typically, an identified model with the necessary model fidelity will not be derived in the first iteration. Thus, additional iteration steps might have to be carried out to obtain a suitable model.

Therefore, the last step is the possible iteration, i.e. the repeated conduction of measurements and evaluation of the measurements until a model meeting the imposed requirements has been found. One often has to conduct initial experiments, which allow to prepare and conduct the main experiments with better suited parameters or methods.

1.3 Taxonomy of Identification Methods and Their Treatment in This Book

According to the definition of identification as presented in the last section, the different identification methods can be classified according to the following criteria:

- Class of mathematical model
- Class of employed test signals
- Calculation of error between process and model

It has proven practical to also include the following two criteria:

- Execution of experiment and evaluation (online, offline)
- Employed algorithm for data processing

Mathematical models which describe the dynamic behavior of processes can be given either as functions relating the input and the output or as functions relating internal states. They can furthermore be set up as analytical models in the form of mathematical equations or as tables or characteristic curves. In the former case, the
parameters of the model are explicitly included in the equation, in the latter case, they are not. Since the parameters of the system play a dominant role in identification, mathematical models shall first and foremost be classified by the model type as:

- Parametric models (i.e. models with structure and finite number of parameters)
- Non-parametric models (i.e. models without specific structure and infinite number of parameters)

**Parametric models** are equations, which explicitly contain the process parameters. Examples are differential equations or transfer functions given as an algebraic expression. **Non-parametric models** provide a relation between a certain input and the corresponding response by means of a table or sampled characteristic curve. Examples are impulse responses, step responses, or frequency responses presented in tabular or graphical form. They implicitly contain the system parameters. Although one could understand the functional values of a step response as “parameters”, one would however need an infinite number of parameters to fully describe the dynamic behavior in this case. Consequently, the resulting model would be of infinite dimension. In this book, parametric models are thus understood as models with a finite number of parameters. Both classes of models can be sub-divided by the type of input and output signals as continuous-time models or discrete-time models.

The **input signals** respectively **test signals** can be deterministic (analytically describable) stochastic (random), or pseudo-stochastic (deterministic, but with properties close to stochastic signals).

As a measure for the **error** between model and process, one can choose between (see Fig. 1.8) the following errors:

- Input error
- Output error
- Generalized equation error
Because of mathematical reasons, typically those errors are preferred, which depend linearly on the process parameters. Thus, one uses the output error if e.g. impulse responses are used as models and the generalized equation error if e.g. differential equations, difference equations, or transfer functions are employed. However, also output errors are used in the last case.

If digital computers are utilized for the identification, then one differentiates between two types of coupling between process and computer, see Fig. 1.9:

- Offline (indirect coupling)
- Online (direct coupling)

For the offline identification, the measured data are first stored (e.g. data storage) and are later transferred to the computer utilized for data evaluation and are processed there. The online identification is performed parallelly to the experiment. The com-
puter is coupled with the process and the data points are operated on as they become available.

The identification with digital computers also allows to discern the identification according to the type of algorithm employed:

- Batch processing
- Real-time processing

In case of **batch processing**, the previously stored measurements will be processed in one shot, which is typically the case for offline applications. If the data are processed immediately after they become available, then one speaks of **real-time processing**, which necessitates a direct coupling between the computer and the process, see Fig. 1.9. Another feature is the processing of the data. Here, one can discern:

- Non-recursive processing
- Recursive processing

The **non-recursive methods** determine the model from the previously stored measurements and are thus a method of choice for offline processing only. On the contrary, the **recursive method** updates the model as each measurement becomes available. Hence, the new measurement is always used to improve the model derived in the previous step. The old measurements do not need to be stored. This is the typical approach for real-time processing and is called **real-time identification**. As not only the parameters, but also a measure of their accuracy (e.g. variance) can be calculated online, one can also think about running the measurement until a certain accuracy of the parameter estimates has been achieved (Åström and Eykhoff, 1971).

Finally, the non-recursive method can further be subdivided into:

- Direct processing
- Iterative processing

The **direct processing** determines the model in one pass. The **iterative processing** determines the model step-wise. Thus, iteration cycles are emerging and the data must be processed multiple times.

### 1.4 Overview of Identification Methods

The most important identification methods shall be described shortly. Table 1.2 compares their most prominent properties. A summary of the important advantages and disadvantages of the individual methods can be found in Sect. 23.4.

#### 1.4.1 Non-Parametric Models

*Frequency response measurements* with periodic test signals allow the direct determination of discrete points of the frequency response characteristics for linear processes. The orthogonal correlation method has proven very effective for this task and is included in all frequency response measurement units. The necessary measurement
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Determination of Characteristic Values</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>-</td>
<td>must be very large</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Average</td>
<td>• Rough model</td>
</tr>
<tr>
<td>Fourier Analysis</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>-</td>
<td>must be large</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>Average</td>
<td>• Validation of theoretically derived models</td>
</tr>
<tr>
<td>Frequency Response Measurement</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>-</td>
<td>average</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>Very good</td>
<td>• Validation of theoretically derived models, Design of classical (linear) controllers</td>
</tr>
<tr>
<td>Correlation Analysis</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>-</td>
<td>can be small</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Good</td>
<td>• Determination of signal relations, Determination of time delays</td>
</tr>
<tr>
<td>Model Adjustment</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>must be large</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>Average</td>
<td>• Parameterization of models</td>
</tr>
<tr>
<td>Parameter Estimation</td>
<td>b_0 + b_1 s</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>can be small</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Good</td>
<td>• Design of adaptive controllers, Adaptive controllers, Fault detection</td>
</tr>
</tbody>
</table>

✓ = Applicable; - = Not applicable
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Iterative Optimization</td>
<td><img src="image1" alt="Diagram" /></td>
<td>-</td>
<td>✓</td>
<td>can be small</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>Bad to Very Good • Design of nonlinear controllers • Fault detection • Parameterization of models</td>
</tr>
<tr>
<td>Extended Kalman</td>
<td><img src="image2" alt="Diagram" /></td>
<td>✓</td>
<td>✓</td>
<td>average</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Average • Combined state and parameter estimation (e.g. no measurement of intermediate quantities) • State estimation of nonlinear systems</td>
</tr>
<tr>
<td>Subspace Methods</td>
<td><img src="image3" alt="Diagram" /></td>
<td>✓</td>
<td>-</td>
<td>can be small</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>Good • Used for Modal Analysis</td>
</tr>
<tr>
<td>Neural Network</td>
<td><img src="image4" alt="Diagram" /></td>
<td>✓</td>
<td>✓</td>
<td>can be small</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>Average • Design of nonlinear controllers • Fault detection • Modeling with little to no knowledge about the underlying process physics</td>
</tr>
</tbody>
</table>

✓ = Applicable; (✓) = Possible, but not well suited; - = Not applicable
time is long if multiple frequencies shall be evaluated, but the resulting accuracy is very high. These methods are covered in this book in Chap. 5.

Fourier analysis is used to identify the frequency response from step or impulse responses for linear processes. It is a simple method with relatively small computational expense and short measurement time, but at the same time is only suitable for processes with good signal-to-noise ratios. A full chapter is devoted to Fourier analysis, see Chap. 3.

Correlation analysis is carried out in the time domain and works with continuous-time as well as discrete-time signals for linear processes. Admissible input signals are both stochastic and periodic signals. The method is also suitable for processes with bad signal-to-noise ratios. The resulting models are correlation functions or in special cases impulse responses for linear processes. In general, the method has a small computational expense. Correlation analysis is discussed in detail in Chap. 6 for the continuous-time case and Chap. 7 for the discrete-time case.

For all non-parametric identification techniques, it must only be ensured a priori that the process can be linearized. A certain model structure does not have to be assumed, what makes these methods very well suited for both lumped as well as distributed parameter systems with any degree of complexity. They are favored for the validation of theoretical models derived from theoretical considerations. Non-parametric models are favored since in this particular area of application, one is not interested in making any a priori assumptions about the model structure.

1.4.2 Parametric Models

For these methods, a dedicated model structure must be assumed. If assumed properly, more precise results are expected due to the larger amount of a priori knowledge.

The most simple method is the determination of characteristic values. Based on measured step or impulse responses, characteristic values, such as the delay time, are determined. With the aid of tables and diagrams, the parameters of simple models can then be calculated. These methods are only suitable for simple processes and small disturbances. They can however be a good starting point for a fast and simple initial system examination to determine e.g. approximate time constants, which allow the correct choice of the sample time for the subsequent application of more elaborate methods of system identification. The determination of characteristic values is discussed in Chap. 2.

Model adjustment methods were originally developed in connection with analog computers. However, they have lost most of their appeal in favor of parameter estimation methods.

Parameter estimation methods are based on difference or differential equations of arbitrary order and dead time. The methods are based on the minimization of certain error signals by means of statistical regression methods and have been complemented with special methods for dynamic systems. They can deal with an arbitrary excitation and small signal-to-noise ratios, can be utilized for manifold applications, work also in closed-loop, and can be extended to non-linear systems. A main focus of the book is placed on these parameter estimation methods. They are discussed e.g. in
Chap. 8, where static non-linearities are treated, Chap. 9, which discusses discrete-time dynamic systems, and Chap. 15, which discusses the application of parameter estimation methods to continuous-time dynamic systems.

Iterative optimization methods have been separated from the previously mentioned parameter estimation methods as these iterative optimization methods can deal with non-linear systems easily at the price of employing non-linear optimization techniques along with all the respective disadvantages.

Subspace-based methods have been used successfully in the area of modal analysis, but have also been applied to other areas of application, where parameters must be estimated. They are discussed in Chap. 16.

Also, neural networks as universal approximators have been applied to experimental system modeling. They often allow to model processes with little to no knowledge of the physics governing the process. Their main disadvantage is the fact that for most neural networks, the net parameters can hardly be interpreted in a physical sense, making it difficult to understand the results of the modeling process. However, local linear neural nets mitigate these disadvantages. Neural nets are discussed in detail in Chap. 20.

The Kalman filter is not used for parameter estimation, but is rather used for state estimation of dynamic systems. Some authors suggest to use the Kalman filter to smoothen the measurements as part of applying parameter estimation methods. A more general framework, the extended Kalman Filter allows the parallel estimation of states and parameters of both linear and non-linear systems. Its use for parameter estimation is reported in many citations. Chapter 21 will present the derivation of the Kalman filter and the extended Kalman filter and outline the advantages and disadvantages of the use of the extended Kalman filter for parameter estimation.

1.4.3 Signal Analysis

The signal analysis methods shown in Table 1.2 are employed to obtain parametric or non-parametric models of signals. Often, they are used to determine the frequency content of signals. The methods differ in many aspects.

A first distinction can be made depending on whether the method is used for periodic, deterministic signals or for stochastic signals. Also, not all methods are suited for time-variant signals, which in this context shall refer to signals, whose parameters (e.g. frequency content) change over time. There are methods available that work entirely in the time domain and others that analyze the signal in the frequency domain.

Not all methods are capable of making explicit statements on the presence or absence of single spectral components, i.e. oscillations at a certain single frequency, thus this capability represents another distinguishing feature. While many methods are capable of detecting periodic components in a signal, many methods can still not make a statement whether the recorded section of the signal is in itself periodic or not. Also, not all methods can determine the amplitude and the phase of the periodic signal components. Some methods can only determine the amplitude and some
## Table 1.2: Overview of the most prominent signal analysis methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Periodic Signal</th>
<th>Stochastic Signal</th>
<th>Time-Variant</th>
<th>Time Domain</th>
<th>Frequency Domain</th>
<th>Single Frequency Detection</th>
<th>Periodicity</th>
<th>Amplitude</th>
<th>Phase</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandpass Filtering</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>(✓)</td>
<td>(✓)</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>• Accuracy depends on filter passband</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Old values do not need to be stored</td>
</tr>
<tr>
<td>Fourier Analysis</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>• Classical and easy to understand tool</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Fast Fourier Transform existst in many implementations</td>
</tr>
<tr>
<td>Parametric Spectral Estimation</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>• Implementations available that do not suffer from windowing</td>
</tr>
<tr>
<td>Correlation Analysis</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
<td>• Time domain method that allows to detect periodicity of signals and determine the period length</td>
</tr>
<tr>
<td>Spectrum Analysis</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>• Can use FFT as for Fourier Analysis</td>
</tr>
<tr>
<td>ARMA Parameter Estimation</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>• Provides coefficients of a form filter that generates the signal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Typically slow convergence of the parameter estimation method</td>
</tr>
<tr>
<td>Short Time Fourier Transform</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>• Blockwise application of the Fourier analysis</td>
</tr>
<tr>
<td>Wavelet Analysis</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>• Well suited for signals with steep transients, e.g. rectangular waves and pulse type oscillations</td>
</tr>
</tbody>
</table>

✓ = Applicable; - = Not Applicable
methods can neither determine the amplitude nor the phase without a subsequent analysis of the results delivered by the signal analysis method.

**Bandpass filtering** uses a bank of bandpass filters to analyze different frequency bands. The biggest advantage of this setup is that past values do not need to be stored. The frequency resolution depends strongly on the width of the filter passband.

**Fourier analysis** is a classical tool to analyze the frequency content of signals and is treated in detail in Chap. 3. The biggest advantage of this method is the fact that many commercial as well as non-commercial implementations of the algorithms exist.

**Parametric spectral estimation** methods can provide signal models as a form filter shaping white noise. They can also decompose a signal into a sum of sinusoidal oscillations. These methods are much less sensitive to the choice of the signal length than e.g. the Fourier analysis, where the sampling interval length typically has to be an integer multiple of the period length. These methods are discussed in Sect. 9.2.

**Correlation analysis** is discussed in detail in Chaps. 6 and 7. It is based on the correlation of a time signal with a time-shifted version of the same signal and is extremely well suited to determine whether a time signal is truly periodic and determine its period length.

**Spectrum analysis** examines the Fourier transform of the auto-correlation function, while the **ARMA parameter estimation** determines the coefficients of an ARMA form filter that generates the stochastic content of the signal. This will be presented in Sect. 9.4.2.

Finally, methods have been developed that allow a joint time-frequency analysis and can be used to check for changes in the signal properties. The **short time Fourier transform** applies the Fourier transform to small blocks of the recorded signals. The **wavelet analysis** calculates the correlation of the signal with a mother wavelet that is shifted and/or scaled in time. Both methods are presented in Chap. 3.

### 1.5 Excitation Signals

For identification purposes, one can supply the system under investigation either with the operational input signals or with artificially created signals, so-called test signals. Such test signals must in particular be applied, if the operational signals do not excite the process sufficiently (e.g. due to small amplitudes, non-stationarity, adverse frequency spectrum), which is often the case in practical applications. The favorable signals typically satisfy the following criteria:

- Simple and reproducible generation of the test signal with or without signal generator
- Simple mathematical description of the signal and its properties for the corresponding identification method
- Realizable with the given actuators
- Applicable to the process
- Good excitation of the interesting system dynamics
Often, one cannot influence the input \( u_2(t) \) that is directly acting on the subprocess \( P_2 \), which is to be identified. The input can only by influenced by means of the preceding subprocess \( P_1 \) (e.g. actuator) and its input \( u_1(t) \), see Fig. 1.11. If \( u_2(t) \) can be measured, the subprocess \( P_2 \) can be identified directly, if the identification method is applicable for the properties of \( u_2(t) \). Is the method applicable for a special test signal \( u_1(t) \) only, then one has to identify the entire process \( P \) and the sub-process \( P_1 \) and calculate \( P_2 \), which for linear systems is given as

\[
G_{P2}(s) = \frac{G_P(s)}{G_{P1}(s)},
\]

(1.5.1)

where the \( G(s) \) are the individual transfer functions.
1.6 Special Application Problems

There are a couple of application problems, which shall be listed here shortly to sensitize the reader to these issues. They will be treated in more detail in later chapters.

1.6.1 Noise at the Input

So far, it was assumed that the disturbances acting on the process can be combined into a single additive disturbance at the output, $y_z$. If the measurement is disturbed by a disturbance $\xi_y(t)$, see Fig. 1.12, then this can be treated together with the disturbance $y_z(t)$ and thus does not pose a significant problem. More difficult is the treatment of a disturbed input signal $u(t)$, being counterfeit by $\xi_u(t)$. This is denoted as errors in variables, see Sect. 23.6. One approach to solve this problem is the method of total least squares (TLS) or the principal component analysis (PCA), see Chap. 10.

Proportional acting processes can in general be identified in open-loop. Yet, this is often not possible for processes with integral action as e.g. interfering disturbance signals may be acting on the process such that the output drifts away. Also, the process may not allow a longer open loop operation as the operating point may start to drift. In these cases as well as for unstable processes, one has to identify the process in closed-loop, see Fig. 1.13. If an external signal such as the setpoint is measurable, the process can be identified with correlation or parameter estimation methods. If there is no measurable external signal acting on the process (e.g. regulator settings with constant setpoint) and the only excitation of the process is by $y_z(t)$, then one is restricted in the applicable methods as well as the controller structure. Chapter 13 discusses some aspects that are proprietary to identification in closed-loop.

1.6.2 Identification of Systems with Multiple Inputs or Outputs

For linear systems with multiple input and/or output signals, see Fig. 1.14, one can also employ the identification methods for SISO processes presented in this book. For a system with one input and $r$ outputs and one test signal, one can obtain $r$
input/output models by applying the identification method \( r \) times to the individual input/output combinations, see Fig. 1.14. A similar approach can be pursued for systems with \( r \) inputs and one output (MISO). One can excite one input after the other or one can excite all inputs at the same time with non-correlated input signals. The resulting model does not have to be the minimum realizable model, though.

For a system with multiple inputs and outputs (MIMO), one has three options: One can excite one input after the other and evaluate all outputs simultaneously, or one can excite all inputs at the same time and evaluate one output after the other, or one can excite all inputs simultaneously and also evaluate all outputs at the same time. If a model for the input/output behavior is sufficient, then one can successfully apply the SISO system identification methods. If however, one has \( p \) inputs which are excited simultaneously and \( r \) outputs, one should resort to methods specifically tailored to the identification of MIMO systems, as the assumed model structure plays an important role here. Parameter estimation of MIMO systems is discussed in Chap. 17.

1.7 Areas of Application

As already mentioned, the application of the resulting model has significant influence on the choice of the model classes, the required model fidelity, the identification method and the required identification hardware and software. Therefore, some sample areas of application shall be sketched in the following.

1.7.1 Gain Increased Knowledge about the Process Behavior

If it proves impossible to determine the static and dynamic behavior by means of theoretical modeling due to a lack of physical insight into the process, one has to resort to experimental modeling. Such complicated cases comprise technical processes
as e.g. furnaces, combustion engines, bio-reactors, biological and economical processes. The choice of the identification method is mainly influenced by the questions whether or not special test signals can be inserted, whether the measurements can be taken continuously or only at discrete points in time, by the number of inputs and outputs, by the signal-to-noise ratio, by the available time for measurements and by the existence of feedback loops. The derived model must typically only be of good/medium fidelity. Often, it is sufficient to apply simple identification methods, however, one also applies parameter estimation methods quite frequently.

1.7.2 Validation of Theoretical Models

Due to the simplifying assumptions and the imprecise knowledge of process parameters, one quite frequently needs to validate a theoretically derived model with experiments conducted at a real process. For a (linear) model given in the form of a transfer function, the measurement of the frequency response provides a good tool to validate the theoretical model. The Bode diagram provides a very transparent representation of the dynamics of the process, such as resonances, the negligence of the higher frequent dynamics, dead time and model order. The major advantage of the frequency response measurement is the fact that no assumptions must be made about the model structure (e.g. model order, dead time,...). The most severe disadvantage is the long measurement time especially for processes with long settling times and the necessary assumption of linearity.

In the presence of mild disturbances only, it may also be sufficient to compare step responses of process and model. This is, of course, very transparent and natural. In the presence of more severe disturbances, however, one has to resort to correlation methods or parameter estimation methods for continuous-time models. The required model fidelity is medium to high.

1.7.3 Tuning of Controller Parameters

The rough tuning of parameters, e.g. for a PID controller, does not necessarily require a detailed model (like Ziegler-Nichols experiment). It is sufficient to determine some characteristic values from the step response measurement. For the fine-tuning however, the model must be much more precise. For this application, parameter estimation methods are favorable, especially for self-tuning digital controllers, see e.g. (Åström and Wittenmark, 1997, 2008; Bobál et al, 2005; O’Dwyer, 2009; Crowe et al, 2005; Isermann et al, 1992). These techniques should gain more momentum in the next decades as the technicians are faced with more and more controllers installed in plants and nowadays more than 50% of all controllers are not commissioned correctly, resulting in slowly oscillating control loops or inferior control performance (Pfeiffer et al, 2009).

1.7.4 Computer-Based Design of Digital Control Algorithms

For the design of model-based control algorithms, for e.g. internal model or predictive controllers or multi-variable controllers, one needs models of relatively high
fidelity. If the control algorithms as well as the design methods are based on para-
metric, discrete-time models, parameter estimation methods, either offline or online,
are the primary choice. For non-linear systems, either parameter estimation methods
or neural nets are suitable (Isermann, 1991).

1.7.5 Adaptive Control Algorithms

If digital adaptive controllers are employed for processes with slowly time-varying
coefficients, parametric discrete-time models are of great benefit since suitable mod-
els can be determined in closed loop and online by means of recursive parameter
estimation methods. By the application of standardized controller design methods,
the controller parameters can be determined easily. However, it is also possible to
employ non-parametric models. This is treated e.g. in the books (Sastry and Bodson,
Adaptive controllers are another important subject due to the same reasons already
stated for the automatic tuning of controller parameters. However, the adaptation
depends very much on the kind of excitation and has to be supervised continuously.

1.7.6 Process Supervision and Fault Detection

If the structure of a process model is known quite accurately from theoretical con-
siderations, one can use continuous-time parameter estimation methods to determine
the model parameters. Changes in the process parameters allow to infer on the pres-
ence of faults in the process. The analysis of the changes also allows to pinpoint the
type of fault, its location and size. This task however imposes high requirements on
the model fidelity. The primary choice are online identification methods with real
time data processing or block processing. For a detailed treatment of this topic, see
e.g. the book by Isermann (2006). Fault detection and diagnosis play an important
role for safety critical systems and in the context of asset management, where all
production equipment will be incorporated into a company wide network and all
equipment will permanently assess its own state of health and request maintenance
service autonomously upon the detection of tiny, incipient faults, which can cause
harmful system behavior or stand-still of the production in the future.

1.7.7 Signal Forecast

For slow processes, such as e.g. furnaces or power plants, one is interested in fore-
casting the effect of the operator intervention by means of a simulation model to
support the operator and enable him/her to judge the effects of his/her intervention.
Typically, recursive online parameter estimation methods are exploited for the task
of deriving a plant model. These methods have also been used to the prediction of
economical markets as described e.g. by (Heij et al, 2007) as well as Box et al (2008).
Table 1.3. Bibliographical list for books on system identification since 1992 with no claim on completeness. ✔=Yes, (√)=Yes, but not covered in depth, C=CD-ROM, D=Diskette, M=MatLab code or toolbox, W=Website. For a reference on books before 1992, see (Isermann, 1992). Realization theory based methods in (Juang, 1994; Juang and Phan, 2006) are sorted as subspace methods.
1.7.8 On-Line Optimization

If the task is to operate a process in its optimal operating point (e.g. for large Diesel ship engines or steam power plants), parameter estimation methods are used to derive an online non-linear dynamic model which then allows to find the optimal operating point by means of mathematical optimization techniques. As the price for energy and production goods, such as crude oil and chemicals, is increasing at a rapid level, it will become more and more important to operate the process as efficiently as possible.

From this variety of examples, one can clearly see the strong influence of the intended application on the choice of the system identification methods. Furthermore, the user is only interested in methods that can be applied to a variety of different problems. Here, parameter estimation methods play an important role since they can easily be modified to not only include linear, time-invariant SISO processes, but also cover non-linear, time-varying, and multi-variable processes.

It has also been illustrated that many of these areas of application of identification techniques will present attractive research and development fields in the future, thus creating a demand for professionals with a good knowledge of system identification. A selection of applications of the methods presented in this book is presented in Chap. 24.

1.8 Bibliographical Overview

The development of system identification has been pushed forward by new developments in diverse areas:

- System theory
- Control engineering
- Signal theory
- Time series analysis
- Measurement engineering
- Numerical mathematics
- Computers and micro-controllers

and Saint-Malo (2009). The *International Federation of Automatic Control* (IFAC) has also devoted the work of the Technical Committee 1.1 to the area of modeling, identification, and signal processing.

Due to the above-mentioned fact that system identification is under research in many different areas, it is difficult to give an overview over all publications that have appeared in this area. However, Table 1.3 tries to provide a list of books that are devoted to system identification, making no claim on completeness. As can be seen from the table many textbooks concentrate on certain areas of system identification.

**Problems**

1.1. **Theoretical Modeling**
Describe the theoretical modeling approach. Which equations can be set up and combined into a model? Which types of differential equations can result? Why is the application of purely theoretical modeling approaches limited?

1.2. **Experimental Modeling**
Describe the experimental modeling approach. What are its advantages and disadvantages?

1.3. **Model Types**
What are white-box, gray-box, and black-box models?

1.4. **Identification**
What are the tasks of the identification?

1.5. **Limitations in Identification**
Which limitations are imposed on a practical identification experiment?

1.6. **Disturbances**
Which typical disturbances are acting on the process? How can their effect be eliminated?

1.7. **Identification**
Which steps have to be taken in the sequence of system identification?

1.8. **Taxonomy of Identification Methods**
According to which features can identification methods be classified?

1.9. **Non-Parametric/Parametric Models**
What is the difference between a non-parametric and a parametric model? Give examples.

1.10. **Areas of Application**
Which identification methods are suitable for validation of theoretical linear models and the design of digital control algorithms.
References

DIN Deutsches Institut für Normung e V (Juli 1998) Begriffe bei Prozessrechensystemen


Identification of Dynamic Systems
An Introduction with Applications
Isermann, R.; Münchhof, M.
2011, XXV, 705 p. With Based on the book: Isermann:
Identifikation dynamischer Systeme 1, 2. With online
files/update., Hardcover
ISBN: 978-3-540-78878-2