1

Introduction

1.1 Introduction to System Identification

In this chapter a brief introduction to linear and nonlinear system identification will be provided. The descriptions are not meant to be detailed or comprehensive. Rather, the aim is to briefly describe the methods from a descriptive point of view so the reader can appreciate the broad development of the methods and the context in which they were introduced. Maths is largely avoided in this first chapter because detailed definitions and descriptions of the models, systems, and identification procedures will be given in the following chapters.

The main theme of the book – methods based around the NARMAX (nonlinear autoregressive moving average model with exogenous inputs) model and related methods – will also be introduced. In particular, the NARMAX philosophy for nonlinear system identification will be briefly described, again with full details given in later chapters, and how this leads into the important problems of frequency response functions for nonlinear systems and models of spatio-temporal systems will be briefly developed.

1.1.1 System Models and Simulation

The concept of a mathematical model is fundamental in many branches of science and engineering. Virtually every system we can think of can be described by a mathematical model. Some diverse examples are illustrated in Figure 1.1. All the systems illustrated in Figure 1.1 can be described by a set of mathematical equations, and this is referred to as the mathematical model of the system. The examples included here show a coal-fired power station, an oil rig, an economic system represented by dealing screens in the stock exchange, a machine vision system (autonomous
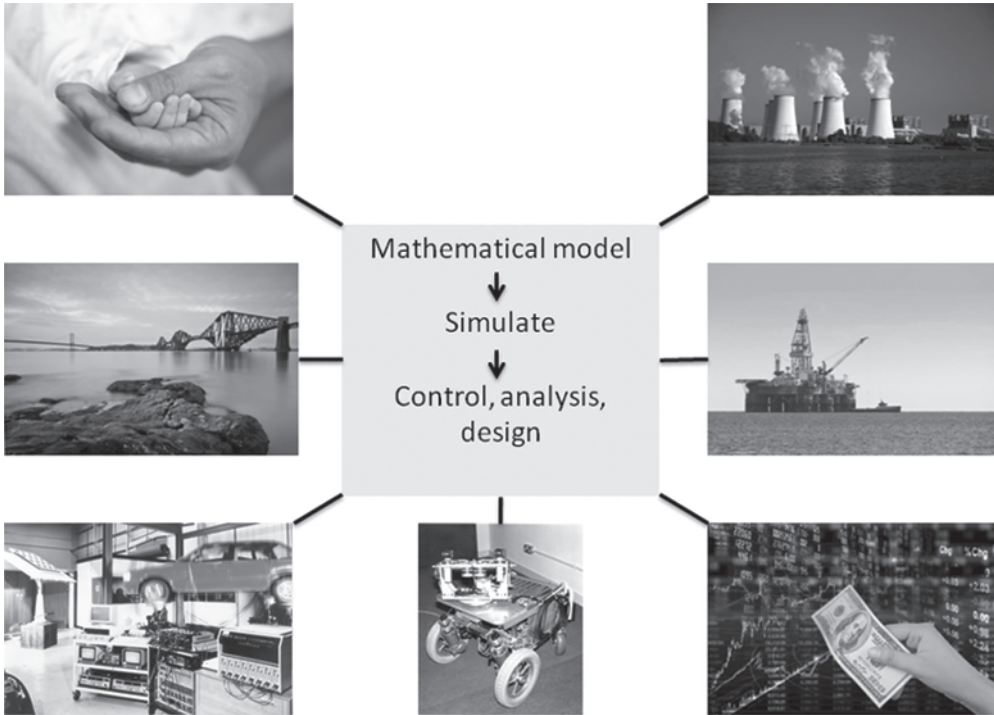
Figure 1.1   Examples of modelling, simulation, and control. Courtesy of dreamstime.com. For a color version of this figure, please refer to the color plates

guided vehicle), a vibrating car, a bridge structure, and a biomedical system. Although each system is made up of quite different components, if each is considered as a system with inputs and outputs that are related by dynamic behaviours then they can all be described by a mathematical model. Surprisingly, all these systems can be represented by just a few basic mathematical operations – such as derivatives and integrals – combined in some appropriate manner with coefficients. The idea of the model is that it describes each system such that the model encodes information about the dynamics of the system. So, for example, a model of the power station would consist of a set of mathematical equations that describe the operation of pulverising the coal, burning it to produce steam, the turbo-alternator, and all the other components that make up this system. Mathematical models are at the heart of analysis, simulation, and design.

Assuming that accurate models of the systems can be built then computers can be programmed to simulate the models, to solve the mathematical equations that represent the system. In this way the computer is programmed to behave like the system. This has numerous advantages: different system designs can be assessed without the expense and delay of physically building the systems, experiments on the computer which would be dangerous on the real system (e.g., nuclear) can be simulated, and information about how the system would respond to different inputs can be acquired. Questions such as 'how does the spacecraft behave if the re-entry angle is changed or one of the rockets fails?', or 'how would the economy respond to a cut in interest rates, would this increase/decrease inflation/unemployment?', and so on, can all be posed and answered. Models therefore are central to the study of dynamical systems.

1.1.2 Systems and Signals

A mathematical model of a system can be used to emulate the system, predict the system response for given inputs, and investigate different design scenarios. However, these objectives can only be achieved if the model of the system is known. The validity of all the simulation, analysis, and design of the system is dependent on the model being an accurate representation of the system. The construction of accurate dynamic models is therefore fundamental to this type of analysis. So how are mathematical models of systems determined?

One way, called analytical modelling, involves breaking the system into component parts and applying the laws of physics and chemistry to each part to slowly build up a description. For example, a resistor can be described by Ohms law, mechanical systems by force and energy balance equations, and heat conduction systems by the laws of thermodynamics, and so on. This process can clearly be very complex, it is time-consuming and may take several man-years, it is problem-dependent, requires a great deal of expertise in many diverse areas of science, and would need to be repeated if any part of the system changed through redesign.

But, returning to the examples of the dynamic systems in Figure 1.1 suggests there is an alternative approach which overcomes most of these problems and which is generally applicable to all systems. Given the mathematical model and the input to a system, the system response can be computed; this is the simulation problem. All the systems in Figure 1.1 produce input and output signals, and if these can be measured it should be possible to work out what the system model must have been. This is the converse to the simulation problem – given measurements of the system inputs and outputs, determine what the mathematical model of the system should be. This is called 'system identification'; it provides the link between systems and signals and is the unifying theme throughout this book. System identification therefore is just a means of measuring the mathematical model of a process.

1.1.3 System Identification

System identification is a method of measuring the mathematical description of a system by processing the observed inputs and outputs of the system. System identification is the complement of the simulation problem. Surely the output signal contains buried within it the dynamics of the mathematical model that produced this signal from the measured input, so how can this information be extracted? System identification provides a principled solution to this problem. Even in ideal conditions this is not easy because the form that the model of the system takes will be unknown, is it linear or nonlinear, how many terms are in the model, what type of terms should be in the model, does the system have a time delay, what type of nonlinearity describes this system, etc.? Yet, if system identification is to be useful, these problems must be resolved. The advantages of system identification are many: it is applicable to all systems, it is often quick, and can be made to track changes in the system. These advantages all suggest that system identification will be a worthwhile study.

1.2 Linear System Identification

Linear systems are defined as systems that satisfy the superposition principle. Linear system identification can be broadly categorised into two approaches; nonparametric and parametric methods. Interest in linear system identification gathered significant momentum from the

1970s onwards, and many new and important results and algorithms were developed (Lee, 1964; Deutsch, 1965; Box and Jenkins, 1970; Himmelblau, 1970; Astrom and Eykoff, 1971; Graupe, 1972; Eykhoff, 1974; Nahi, 1976; Goodwin and Payne, 1977; Ljung and Södeström, 1983; Young, 1984; Norton, 1986; Ljung, 1987; Södeström and Stoica, 1989; Keesman, 2011). Nonparametric methods develop models based typically on the system impulse response or frequency response functions (Papoulis, 1965; Jenkins and Watts, 1968; Eykhoff, 1974; Pintelon and Schoukens, 2001; Bendat and Piersol, 2010). These are usually based on correlation methods and Fourier transforms, respectively, although there are many alternative methods. Special input signals were developed at this time, including multi-level sequences, of which the pseudo-random binary sequence was particularly important (Godfrey, 1993). Pseudo-random sequences could be easily designed and generated and were an ideal sequence to use in experiments on industrial plants to identify linear models. The sequences could be tailored to the process under investigation, so that the power of the input excitation was matched to the bandwidth of the process. This had the advantage that the noise-free signal output was maximised and hence the signal-to-noise ratio on the measured output was enhanced. Pseudo-random binary sequences were the best approximation to white noise and this led to important advantages when using cross-correlation to identify the models because if the input was correctly designed, so that the autocorrelation of the input was an impulse at the origin, the Wiener–Hopf equation (Jenkins and Watts, 1968; Priestley, 1981; Bendat and Piersol, 2010) which relates the cross-correlation between the input and output of a system to the convolution of the system impulse response and the autocorrelation function simplifies so that the cross-correlation becomes directly proportional to the system impulse response. This was a significant result, and the use and development of pseudo-random sequences continued for many years. The other advantage of using a designed input, not just a pseudo-random sequence, was that the input could be measured almost perfectly.

The introduction of the fast Fourier transform (FFT) in 1965 (Jenkins and Watts, 1968) meant that previously slow methods of computing the Fourier transform of a data sequence became much faster and efficient, with increases in speed of orders of magnitude. Linear system identification methods based on the cross and power spectral densities were further developed, following the introduction of the FFT, to provide estimates of the system frequency response. The advantages of these approaches, which replaced the convolution in time with the much simpler algebraic relationships in the Laplace and frequency domains, were offset by the need to window and smooth the spectral estimates to obtain good estimates (Jenkins and Watts, 1968; Bendat and Piersol, 2010). Coherency functions were used to detect poor estimates, and a catalogue of methods was developed based on the frequency response function estimates. This fed into developments in mechanical engineering based on modal analysis (Worden and Tomlinson, 2001), which became established as an important method of analysing and studying vibrations in all kinds of structures.

Parametric methods became popular from the 1970s onwards with an explosion of developments fuelled by the interest at that time in control systems and the development of methods of online process control, and adaptive control including self-tuning algorithms (Wellstead and Zarrop, 1991). These latter methods were all based on a model of the process that could be updated online. Least squares-based methods were developed and the effect of noise on the measurements was studied in depth, resulting in the introduction of algorithms including instrumental variables (Young, 1970), generalised least squares (Clarke, 1967), suboptimal least squares, extended least squares and maximum likelihood (Astrom and Eykhoff, 1971; Eykhoff, 1974). It was realised that data from almost every real system will involve inaccurate

measurements and corruption of the signals by noise. It was shown that if the noise is corre-lated or coloured, biased estimates will be obtained and that even small amounts of correlated noise can result in significantly incorrect models (Astrom and Eykhoff, 1971; Eykhoff, 1974; Goodwin and Payne, 1977; Norton, 1986; Södeström and Stoica, 1989). All the algorithms above therefore were designed to either accommodate the noise or model it explicitly (Clarke, 1967; Young, 1970). Even the offline algorithms were therefore iterative, so that both a model of the process and a model of the noise were identified by operating on the data set several times over until the algorithm converged. Later, in the 1980s, prediction error methods were developed; many of the earlier parameter estimation algorithms were unified under the predic-tion error structure, and elegant proofs of convergence and analysis of the methods were developed (Ljung and Södeström, 1983; Norton, 1986; Ljung, 1987; Södeström and Stoica, 1989). The advantage of the prediction error methods was that they had almost the same asymptotic properties as the maximum likelihood algorithm but, while the probability density function of the residuals had to be known to apply maximum likelihood (which for linear systems could be taken as Gaussian), the prediction error methods optimised a cost function without any knowledge of the density functions (Ljung and Södeström, 1983; Ljung, 1987). This latter point became very important for the development of parameter estimation methods for nonlinear systems, where the signals will almost never be Gaussian and therefore the den-sity functions will rarely be known.

Online or recursive algorithms were also actively developed from the 1970s onwards (Ljung and Södeström, 1983; Young, 1984; Norton, 1986). In contrast to the batch methods described above, where all the data is processed at once, in recursive methods the data is processed over a data window that is moved through the data set. This allows online tracking of slow time variation and is often the basis of adaptive, self-tuning, and many fault-detection algorithms.

The development of linear identification algorithms is still a very active and healthy research field, with many participants from all around the world. This has been encouraged by the ever-increasing need to develop models of systems and the simple fact that system identification is relatively straightforward; it works well most of the time, and can be applied to any system where data can be recorded.

1.3   Nonlinear System Identification

Nonlinear systems are usually defined as any system which is not linear, that is any system that does not satisfy the superposition principle. This contrarian description is very vague but is often necessary because there are so many types of nonlinear systems that it is almost impossible to write down a description that covers all the classes that can exist under the title of 'nonlinear dynamic system'. Authors therefore tend to focus on particular classes of non-linear systems, which can be tightly defined, but which are limited. Historically, system iden-tification for nonlinear systems has developed by focusing on specific classes of system and specific models. The early work was dominated by methods based on the Volterra series, which in the discrete time case can be expressed as

$$
\begin{aligned}
y(k) = h_0 &+ \sum_{m_1=1}^{M} h_1(m_1) u(k-m_1) + \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} h_2(m_1, m_2) u(k-m_1) u(k-m_2) \\
&+ \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} \sum_{m_3=1}^{M} h_3(m_1, m_2, m_3) u(k-m_1) u(k-m_2) u(k-m_3) + \cdots
\end{aligned}
\tag{1.1}
$$

where $u(k), y(k); k = 1, 2, 3 \ldots$ are the measured input and output, respectively, and $h_\ell(m_1, \ldots, m_\ell)$ is the ℓ'th-order Volterra kernel, or ℓ'th-order nonlinear impulse response. The Volterra series is an extension of the linear convolution integral and represents mildly nonlinear systems as a series of multi-summations, or integrals in the continuous time case, of the Volterra kernels and the inputs. Most of the earlier algorithms assumed that just the first two, linear and quadratic, Volterra kernels are present and used special inputs such as Gaussian white noise and correlation methods to identify the two Volterra kernels. Notice that for these early identification methods the input has to be Gaussian and white, which is a severe restriction for many real processes and pre-recorded data sets. These results were later extended to include the first three Volterra kernels, to allow different inputs, and other related developments including the Weiner series. A very important body of work was developed by Wiener, Lee, Bose and colleagues at MIT from the 1940s to the 1960s (Wiener, 1958; Lee, 1964). Much of this work involved developing methods of analysis for nonlinear systems, but important identification algorithms were also introduced including the famous Lee and Schetzen method (1965). The books of Schetzen (1980) and Rugh (1981) describe the many developments based on the work of Volterra and Weiner. While these methods are still actively studied (Marmarelis and Marmarelis, 1978; Doyle *et al*., 2000) as methods of analysis, system identification based on the Volterra (and related Weiner) series is still challenging today. This is because of three basic requirements. First, the number of terms in the Volterra series is unknown at the start of the identification so methods which make assumptions that only the first two or three kernels are present cannot be applied with confidence because there may be many more terms and ignoring these terms will produce incorrect estimates. Second, often special inputs such as Gaussian white noise are required which may not be possible in many real experiments and will not be applicable where data has been pre-recorded. Third, the number of points that need to be identified can be very large. For example, for a system where the first-order Volterra kernel $h_1(m_1)$ is described by say 30 samples, 30×30 points will be required for the second-order kernel $h_2(m_1, m_2)$, $30 \times 30 \times 30$ for the third-order $h_3(m_1, m_2, m_3)$, and so on, and hence the amount of data required to provide good estimates becomes excessively large (Billings, 1980). These numbers can be reduced by exploiting certain symmetries but the requirements are still excessive irrespective of what algorithm is used for the identification. However, the Volterra series is still enormously important as a descriptor of nonlinear systems and as a method of analysis, although this can often be achieved by identifying alternative model forms and then mapping these back to the Volterra model.

Because of the problems of identifying Volterra models, from the late 1970s onwards other model forms were investigated as a basis for system identification for nonlinear systems. Various forms of block-structured nonlinear models were introduced or reintroduced at this time (Billings and Fakhouri, 1978, 1982; Billings, 1980; Haber and Keviczky, 1999). The Hammerstein model consists of a static single-valued nonlinear element followed by a linear dynamic element. The Wiener model is the reverse of this combination, so that the linear element is before the static nonlinear characteristic. The General Model consists of a static linear element sandwiched between two dynamic systems. Other models, such as the S_m, Uryson, etc. models, represent alternative combinations of elements. All these models can be represented by a Volterra series, but in this case the Volterra kernels take on a special form in each case. Identification consists mainly of correlation-based methods, although some parameter estimation methods were also developed. The correlation methods exploited certain properties of these systems which meant that if specific inputs were used, often white Gaussian noise

again, the individual elements could be identified one at a time. This resulted in manageable requirements of data and the individual blocks could sometimes be related to components in the system under study. Methods were developed, based on correlation and separable functions, which could determine which of the block-structured models was appropriate to represent a system (Billings and Fakhouri, 1978, 1982). Many results were introduced and these systems continue to be studied in depth. The problem of course is that these methods are only applicable to a very special form of model in each case and cannot therefore be considered as generic. They make too many assumptions about the form of the model to be fitted, and if little is known about the underlying system then applying a method that assumes a very special model form may not work well. All the above are essentially nonparametric methods of identification for nonlinear systems.

1.4   NARMAX Methods

The NARMAX model was introduced in 1981 as a new representation for a wide class of nonlinear systems (Billings and Leontaritis, 1981; Leontaritis and Billings, 1985; Chen and Billings, 1989). The NARMAX model is defined as

$$
\begin{aligned}
y(k) = F[\,&y(k-1),\, y(k-2),\ldots,\, y(k-n_y),\\
&u(k-d),\, u(k-d-1),\ldots,u(k-d-n_u),\\
&e(k-1),\, e(k-2),\ldots,e(k-n_e)] + e(k)
\end{aligned}
\tag{1.2}
$$

where $y(k)$, $u(k)$, and $e(k)$ are the system output, input, and noise sequences, respectively; n_y, n_u, and n_e are the maximum lags for the system output, input, and noise; $F[\cdot]$ is some nonlinear function, and d is a time delay typically set to $d=1$. The model is essentially an expansion of past inputs, outputs, and noise terms. The exact form of the model and the class of systems that can be represented by this model will be discussed in Chapter 2. However, the essence of the NARMAX model is that past outputs are included in the expansion. The importance of this can be explained by considering linear FIR (finite impulse response) and IIR (infinite impulse response) filters. The FIR filter

$$
y(k) = b_1 u(k-1) + b_2 u(k-2) + \cdots + b_{nb} u(k-nb)
\tag{1.3}
$$

expands the system response in terms of past inputs only. The IIR filter

$$
y(k) + a_1 y(k-1) + \cdots + a_{na} y(k-na) = b_1 u(k-1) + \cdots + b_{nb} u(k-nb)
\tag{1.4}
$$

expands the response in terms of past inputs and outputs, where na and nb represent the model orders. So, for a simple linear system, an FIR filter may typically need 50 weights ($nb=50$) whereas the IIR filter would need maybe 4 ($na=nb=2$), simply because the information in the many past inputs expanded as an FIR filter can be captured by just a few output lagged terms in an IIR filter. The trade-off is that the IIR filter can be more difficult to estimate, but it is far more concise. For nonlinear systems the Volterra series expands the current output as a series in terms of past inputs only. In the nonlinear case this can lead to an explosion in the number of terms to be estimated. It is easy to suggest nonlinear examples where the model inherently has nonlinear output terms, like the Duffing or Van der Pol models (Nayfeh and Mook, 1979; Pearson, 1999), where the output terms in these models will inevitably create a very long

Volterra series. NARMAX, however, can capture these effects easily because nonlinear lagged output terms are allowed. This makes the identification easier because fewer terms are required to represent systems, but it also means that noise on the output has to be taken into account when estimating the model coefficients. The Volterra, block-structured models, and many neural network architectures can all be considered as subsets of the NARMAX model. Since NARMAX was introduced, by proving what class of nonlinear systems can be represented by this model, many results and algorithms have been derived based around this description. Most of the early work was based on polynomial expansions of the NARMAX model. These are still the most popular methods today, but other more complex forms based on wavelets and other expansions have been introduced to represent severely nonlinear and highly complex nonlinear systems. A significant proportion of nonlinear systems can be represented by a NARMAX model, including systems with exotic behaviours such as chaos, bifurcations, and sub-harmonics.

1.5   The NARMAX Philosophy

While NARMAX started as the name of a model, it has now developed into a philosophy of nonlinear system identification (Billings and Tsang, 1989; Billings and Chen, 1992). The NARMAX approach consists of several steps:

Structure detection	Which terms are in the model?
Parameter estimation	What are the model coefficients?
Model validation	Is the model unbiased and correct?
Prediction	What is the output at some future time?
Analysis	What are the dynamical properties of the system?

Structure detection forms the most fundamental part of NARMAX. In linear parameter estimation it is relatively easy to determine the model order. Often models of order one, two, three, and so on are estimated and this is quick and efficient. The models are then validated and compared to find which is the simplest model that can adequately represent the system. This process works well because, assuming a pulse transfer function representation, every increase in model order only increases the number of unknown parameters by two – one extra coefficient for the numerator and the denominator. Over-fitted models are easily detected by pole zero cancellations and other methods.

But this naïve approach does not easily carry over to the nonlinear case. For example, a NARMAX model which consists of one lagged input and one lagged output term, three lagged noise terms, expanded as a cubic polynomial, would consist of 56 possible candidate terms. This number of candidate terms arises because the expansion by definition includes all possible combinations within the cubic expansion. Naïvely proceeding to estimate a model which includes all these terms and then pruning will cause numerical and computational problems and should always be avoided. However, often only a few terms are important in the model. Structure detection, which aims to select terms one at a time, is therefore critically important. This makes sense from an intuitive perspective – build the model by putting in the most important or significant term first, then the next most significant term, and so on, and stop when the model is adequate, it is numerically efficient and sound, and most important of all leads to simple parsimonious models that can be related to the underlying system.

These objectives can easily be achieved by using the orthogonal least squares (OLS) algorithm and its derivatives to select the NARMAX model terms one at a time (Korenberg *et al*., 1988; Billings *et al*., 1989; Billings and Chen, 1998). This approach can be adopted for many different model forms and expansions, and is described in Chapter 3.

These ideas can also be adapted for pattern recognition and feature selection with the advantage that the features are revealed as basis functions that are easily related back to the original problem (Wei and Billings, 2007). The basis vectors are not potentially functions of all the initial features as is the case in principal component analysis, which then destroys easy interpretation of the results.

The philosophy of NARMAX therefore relates to finding the model structure or fitting the simplest model so that the underlying rule is elucidated. Building up the model, term by term, has many benefits not least because if the underlying system is linear, NARMAX methods should just fit a linear model and stop when this model is a good representation of the system. It would be completely wrong to fit a nonlinear model to represent a linear system. For example, the stability of linear systems is well known and is applicable for any input. This does not apply to nonlinear systems. Over-fitting nonlinear systems, by using either excessive time lags or excessive nonlinear function approximations, not only induces numerical problems but can also introduce additional unwanted dynamic behaviours and disguises rather than reveals the relationships that describe the system.

1.6 What is System Identification For?

The fundamental concept of structure detection, that is core to NARMAX methods, naturally leads into a discussion of what system identification is for. Very broadly, this can be divided into two aims.

The first involves approximation, where the key aim is to develop a model that approximates the data set such that good predictions can be made. There are many applications where this approach is appropriate, for example in time series prediction of the weather, stock prices, speech, target tracking, pattern classification, etc. In such applications the form of the model is not that important. The objective is to find an approximation scheme which produces the minimum prediction errors. Fuzzy logic, neural networks, and derivatives of these including Bayesian methods naturally solve these types of problems easily and well (Miller *et al*., 1990; Chen and Billings, 1992; Bishop, 1995; Haykin, 1999; Liu, 2001; Nelles, 2001). The approximation properties of these approaches are usually quoted based on the Weierstrass theorem, which of course equally applies to many other model forms. Naturally, users of these methods focus on the mean-squared-error properties of the fitted model, perhaps over estimation and test sets.

A second objective of system identification, which includes the first objective as a subset, involves much more than just finding a model to achieve the best mean-squared errors. This second aim is why the NARMAX philosophy was developed and is linked to the idea of finding the simplest model structure. The aim here is to develop models that reproduce the dynamic characteristics of the underlying system, to find the simplest possible model, and if possible to relate this to components and behaviours of the system under study. Science and engineering are about understanding systems, breaking complex behaviours down into simpler behaviours that can be understood, manipulated, and exploited. The core aim of this second approach to

identification is therefore, wherever possible, to identify, reveal, and analyse the rule that represents the system. So, if the system can be represented by a simple first-order dynamic system with a cubic nonlinear term in the input this should be revealed by the system identification. Take, for example, two different oil rigs, which are similar but of a different size and operate in different ocean depths and sea states. If the underlying hydrodynamic characteristics which describe the action of the waves on the platform legs and the surge of the platform follow the same scientific law, then the identified models should reveal this (Worden *et al.*, 1994; Swain *et al.*, 1998). That is, we would expect the core model characteristics to be the same even though the parameter values could be different. Therefore, a very important aim is to find the rule so that this can be analysed and understood. Gross approximation to the data is not sufficient in these cases, finding the best model structure is. Ideally, we want to be able to write the identified model down and to relate the terms and characteristics of the model to the system. These aims relate to the understanding of systems, breaking complex behaviours down into simpler behaviours that can be simulated, analysed, and understood. These objectives are relevant to model simulation and control systems design, but increasingly to applications in medicine, neuroscience, and the life sciences. Here the aim is to identify models, often nonlinear, that can be used to understand the basic mechanisms of how these systems operate and behave so that we can manipulate and utilise them.

These arguments also carry over to the requirement to fit models of the system and of the noise. Noise models are important to ensure that the estimated model of the system is unbiased and not just a model of one data set, but noise models are also highly informative. Noise models reveal what is unpredictable from the input, and they indicate the level and confidence that can be placed in any prediction or simulation of the system output.

NARMAX started off as the name of a model class but has now become a generic term for identification methods that aim to model systems in the simplest possible way. Model validation is a critical part of NARMAX modelling and goes far beyond just comparing mean-squared errors. One of the basic approaches involves testing whether there is anything predictable left in the residuals (Billings and Voon, 1986; Billings and Zhu, 1995). The aim is to find the simplest possible model that satisfies this condition. The idea is that if the models of the system and of the noise are adequate, then all the information in the data set should be captured in the model, and the remainder – the final residuals – should be unpredictable from all past inputs and outputs. This is statistical validation and can be applied to any model form and any fitting algorithm. Qualitative validation is also used to develop NARMAX estimation procedures that reproduce the dynamic invariants of the systems. Models that are developed based on term selection to obtain the simplest possible model have been shown to reproduce attractors and dynamic invariants that are topologically closer to the properties of the underlying system dynamics than over-fitted models (Aguirre and Billings, 1995a, b). This links back to the desire to be able to relate the models to the underlying system and to use the models to understand basic behaviours and processes not just to approximate a data set.

NARMAX modelling is a process that can involve feedback in the model-fitting process. As an example, if the initial library of terms that are used to search for the correct model terms is not large enough, then the algorithms will be unable to find the appropriate model. But, applying model validation methods should reveal that terms are missing from the model, and in some instances can suggest what type of terms are missing. The estimation process can then be restarted by including a wider range or different types of model terms. Only when the structure detection and all the validation procedures are satisfied is the model accepted as a

good representation of the system. Just using mean-squared errors is often uninformative and can lead to fitting to the noise, and in the worse case models that are little more than lookup tables.

1.7   Frequency Response of Nonlinear Systems

In the analysis of linear systems a combined time and frequency domain analysis is ubiquitous. Frequency domain methods are core in control system design, vibrations, acoustics, communications, and in almost every branch of science. However, an inspection of the nonlinear system identification literature over the last 20 years or so shows that mainly time domain methods have been developed. Neural networks, fuzzy logic, Bayesian algorithms are all based solely in the time domain and no information about frequency response is supplied. Linear methods would suggest that this is a gross oversight and NARMAX methods have been developed in both time and frequency.

Early methods of computing the generalised frequency response functions (GFRFs) – these are generalisations of the linear frequency response function – were based on the Fourier transform of the Volterra series and hence suffered from all the disadvantages including the need for very long data sets, unrealistic assumptions about the systems, and specialised inputs. However, all these problems can be avoided by mapping identified NARMAX models directly into the GFRFs (Billings and Tsang, 1989; Peyton-Jones and Billings, 1989). This means that the GFRFs can be written down and, importantly, that the effects in frequency can be related back to specific time domain model terms and vice versa. This links back to the importance of finding the simplest model structure and relating that model and its properties to the underlying system characteristics. The linear case can be used to illustrate this point. For linear systems we might identify a state-space model, a weighting or impulse sequence, a pulse transfer function, or several other model forms. When the system is linear all these models are related and any one can readily be transformed into another. If each of these different model forms were identified for a particular system, if the models are unbiased and correct, they should all have exactly the same frequency response. In addition, just looking at time domain behaviours does not always reveal invariant characteristics which are so important in the scientific understanding of basic behaviours in any system. So, even if a correct linear model has been identified, obviously simulating this model with different inputs (maybe a random input and a swept sine) does not easily reveal properties of that system by visual inspection. But if the system is of second order, the frequency response in every case should show one resonance; this can be related to specific terms in the system model and hence back to the system under study, and shows a core invariant system behaviour.

The same argument holds for nonlinear dynamic systems but now the story is more complex. First, many different types of models could be fitted to a data set from a nonlinear system – Volterra, NARMAX, nonlinear state-space, neural networks, etc. But it is often virtually impossible to map from one model to another and, as in the linear case, just looking at properties in the time domain only reveals half the picture. This is why we map NARMAX models to the GFRFs, because this reveals core invariant behaviours that can usually be interpreted in a very revealing manner. Because this is a mapping, each GFRF can be generated one at a time and even if there are a large number it is easy to evaluate which are important and when to stop. Core frequency response behaviours, which are essentially extensions of the concept of

resonance, can then be identified and related back to the behaviour and properties of the underlying system. This process is relatively easy even for complex systems, has been extended to severely nonlinear systems with sub-harmonics and, while the potentially large number of GFRFs may at first appear to be a problem, this can be turned around and used as a great benefit. For example, in the design of a totally new class of filters called energy transfer filters. Frequency domain analysis is therefore core to the NARMAX philosophy and is discussed in Chapters 6 and 7.

1.8   Continuous-Time, Severely Nonlinear, and Time-Varying Models and Systems

The vast majority of system identification methods, certainly for nonlinear systems, are based on discrete time models. This is natural because data collection inevitably involves data sampling, so that the discrete domain is the natural choice. But there are situations where a continuous-time model would be preferable. Continuous-time models are often simpler in structure than the discrete counterpart. For example, a second-order derivative term in continuous time would involve at least three and often more, depending on the approximation scheme, terms in discrete time. Continuous-time models are also independent of the sample rate.

The established literature on most systems and processes is almost always based on continuous-time integro-differential equations. So that, if the identification involves a study of a system that has been analysed before using different modelling approaches such as analytical modelling using the basic laws of science, then an identified continuous-time model can more easily be compared to previous models. In the modelling of the magnetosphere and space weather (see the case studies in Chapter 14 for a specific example), there is a considerable body of analytical modelling work developed by physicists over many years. If nonlinear continuous-time models can be identified then these can be compared to the previous work and indeed the analytical models can be used to prime the model structure selection (Balikhin *et al.*, 2001). Model validation can also be used to validate existing physically derived models and NARMAX methods can be used to find missing model terms and to analyse these models in the frequency domain. This is why we both study the estimation of the structure – that is, what model terms to include – and estimate the parameters in complex nonlinear differential equation models. NARMAX methods can be extended to solve these problems, often without the need to differentiate data which always increases noise considerably.

Severely nonlinear systems that exhibit sub-harmonics are also studied. These results are developed following the philosophy of finding the simplest possible model and because sub-harmonics is a frequency domain behaviour, developing algorithms that allow the user to see the properties in the frequency domain is important (Li and Billings, 2005). These algorithms allow NARMAX to be applied to model very exotic and complex dynamic behaviours.

Time-varying systems have been extensively studied based on classical LMS, recursive least squares, and Kalman filter-based algorithms. But most of the existing methods only work for slow time variation. However, by using a new wavelet expansion-based approach, NARMAX algorithms have been developed to track rapid time changes and movements and to map these to the frequency domain where invariant characteristics can be tracked – for EEG analysis, for example. These problems are discussed in detail in Chapters 9, 10, and 11.

1.9   Spatio-temporal Systems

Spatio-temporal systems are systems that evolve over both space and time (Hoyle, 2006). Purely temporal systems involve measurements of a variable over time. There are also examples where measurements at one spatial location, for example an electrophysiological probe in the brain, or a flow monitor in a river, also produce a temporal signal. But both these examples are strictly spatio-temporal systems. That is, the dynamics at each spatial location may depend, in a nonlinear dynamic way, both on what happened back in time and what happened at other spatial locations back in time. There are many applications of such systems, for example the dynamics of cells in a dish, the growth of crystals, neuro-images, etc. These are a very important and neglected class of systems, and hence NARMAX methods have been developed to identify several different model classes which can be used to represent spatio-temporal behaviours including cellular automata, coupled map lattices, and nonlinear partial differential equations.

The concept of model structure is even more important for spatio-temporal systems because a model of a system may involve just a few lagged time terms at a few, possibly nonadjacent spatial locations. Grossly approximating the system would therefore be inappropriate, and again the key challenge is to find the model structure which now involves finding the neighbourhood that defines the spatial interactions and the temporal lags. Invariant behaviours are also important in spatio-temporal systems, simply because a model excited with different inputs will produce different patterns that evolve over time. Depending on the choice of inputs, the patterns produced from an identical model could be significantly different when inspected visually. Comparing different models and different patterns to discover the rules of the underlying behaviours is therefore very difficult. That is why the GFRFs for NARMAX models have recently been introduced for spatio-temporal NARMAX models. These problems are discussed in detail in Chapters 12 and 13.

1.10   Using Nonlinear System Identification in Practice and Case Study Examples

While there is a considerable literature on algorithms for nonlinear system identification of all sorts of shapes and forms, there are a relatively small number of users who are expert at applying these methods to real-life systems. Most authors just use simulated examples to illustrate and test their algorithms. Linear parameter estimation and NARMAX models can be studied and thoroughly tested by simulating known models and comparing the initial simulated model coefficients to those identified. This provides a powerful means of evaluating the methods. Neural networks, which are designed to purely approximate systems, produce models that usually contain so many weights or parameters and basic approximating units that the model representation cannot be written down, and maybe conveniently therefore cannot be tested to check the training procedures do indeed identify the exact same model that was used as a simulated test to begin with.

This is why the overall aim of this book is to try to introduce and show the reader how to apply NARMAX methods to real problems. The emphasis therefore is on describing the methods in a way that is as transparent as possible, deliberately leaving out all the variants of the methods and their complex derivations and properties, all of which are available in the literature.

Hence, in Chapter 14, practical aspects of nonlinear system identification and many case studies are described. The case studies are deliberately taken from a wide range of systems that we have analysed over recent years and range from modelling space weather systems, through to the identification of the visual system of a fruit fly, to the modelling of iceberg flux in Greenland, and many other systems. All the case studies are for real problems where the main objective is to use system identification as a tool to understand the complex system being studied in a way that is revealing, transparent, and as simple as possible.

References

Aguirre, L.A. and Billings, S.A. (1995a) Dynamical effects of over-parameterisation in nonlinear models. *Physica D, 80*, 26–40.

Aguirre, L.A. and Billings, S.A. (1995b) Retrieving dynamical invariants from chaotic data using NARMAX models. *International Journal of Bifurcation and Chaos, 5*, 449–474.

Astrom, K.J. and Eykhoff, P. (1971) System identification—a survey. *Automatica, 7*, 123–162.

Balikhin, M., Boaghe, O.M., Billings, S.A., and Alleyne, H. (2001) Terrestrial magnetosphere as a nonlinear dynamical resonator. *Geophysical Research Letters, 28*, 1123–1126.

Bendat, J.S. and Piersol, A.G. (2010) *Random Data Analysis and Measurement Procedures*, 4th edn. New York: John Wiley & Sons.

Billings, S.A. (1980) Identification of nonlinear systems: a survey. *IEE Proceedings, Pt. D, 127*(6), 272–285.

Billings, S.A. and Chen, S. (1992) Neural networks and system identification. In K. Warwick, G.W. Irwin and K.J. Hunt (eds), *Neural Networks for Systems and Control*. London: Peter Peregrinus Ltd, on behalf of IEE, pp. 181–205.

Billings, S.A. and Chen, S. (1998) The determination of multivariable nonlinear models for dynamic systems using neural networks. In C.T. Leondes (ed.), *Neural Network System Techniques and Applications*. San Diego, CA: Academic Press, pp. 231–278.

Billings, S.A. and Fakhouri, S.Y. (1978) Identification of a class of nonlinear systems using correlation analysis. *IEE Proceedings, Pt. D, 125*, 691–697.

Billings, S.A. and Fakhouri, S.Y. (1982) Identification of systems containing linear dynamic and static nonlinear elements. *Automatica, 18*(1), 15–26.

Billings, S.A. and Leontaritis, I.J. (1981) Identification of nonlinear systems using parametric estimation techniques. *Proceedings of the IEE Conference on Control and its Application*, Warwick, UK, pp. 183–187.

Billings, S.A. and Tsang, K.M. (1989) Spectral analysis for nonlinear systems—Part I: Parametric nonlinear spectral analysis. *Mechanical Systems and Signal Processing, 3*(4), 319–339.

Billings, S.A. and Voon, W.S.F. (1986) Correlation based model validity tests for non-linear models. *International Journal of Control, 44*(1), 235–244.

Billings, S.A. and Zhu, Q.M. (1995) Model validation tests for multivariable nonlinear models including neural networks, *International Journal of Control, 62*, 749–766.

Billings, S.A., Chen, S., and Korenberg, M.J. (1989) Identification of MIMO non-linear systems using a forward regression orthogonal estimator. *International Journal of Control, 49*(6), 2157–2189.

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

Box, G.E.P. and Jenkins, G.M. (1970) *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Chen, S. and Billings, S.A. (1989) Representation of non-linear systems: the NARMAX model. *International Journal of Control, 49*(3), 1013–1032.

Chen, S. and Billings, S.A. (1992) Neural networks for nonlinear dynamic system modelling and identification. *International Journal of Control, 56*(2), 319–346.

Clarke, D.W. (1967) Generalised least squares estimation of parameters of a dynamic model. IFAC Symposium on System Identification, Prague, pp. 1–11.

Deutsch, R. (1965) *Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Doyle, F.J., Pearson, R.K., and Ogunnaike, B.A. (2000) *Identification and Control using Volterra Models*. Berlin: Springer-Verlag.

Eykhoff, P. (1974) *System Identification-Parameter and State Estimation*. New York: John Wiley & Sons.

Godfrey, K. (1993) *Perturbation Signals for System Identification*. Englewood Cliffs, NJ: Prentice-Hall.

Goodwin, G.C. and Payne, R.L. (1977) *Dynamic System Identification*. New York: Academic Press.

Graupe, D. (1972) *Identification of Systems*. New York: Van Nostrand Reinholt.

Haber, R. and Keviczky, L. (1999) *Nonlinear System Identification–Input Output Modelling Approach*, Vols *I and II*. Dordrecht: Kluwer Academic.

Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.

Himmelblau, D.M. (1970) *Process Analysis by Statistical Methods*. New York: John Wiley & Sons.

Hoyle, R. (2006) *Pattern Formation*. Cambridge: Cambridge University Press.

Jenkins, G.M. and Watts, D.G. (1968) *Spectral Analysis and its Applications*. San Francisco, CA: Holden-Day.

Keesman, K.J. (2011) *System Identification an Introduction*. Berlin: Springer-Verlag.

Korenberg, M.J., Billings, S.A., Liu, Y.P., and McIlroy, P.J. (1988) Orthogonal parameter estimation algorithm for nonlinear stochastic systems. *International Journal of Control, 48*, 193–210.

Lee, R.C.K. (1964) *Optimal Estimation, Identification and Control*. Cambridge, MA: MIT Press.

Leontaritis, I.J. and Billings, S.A. (1985) Input–output parametric models for non-linear systems—Part I: Deterministic non-linear systems; Part II: Stochastic non-linear systems. *International Journal of Control, 41*(2), 303–328; 329–344.

Li, L.M. and Billings, S.A. (2005) Discrete time sub harmonic modelling and analysis. *International Journal of Control, 78*, 1265–1284.

Liu, G.P. (2001) *Nonlinear Identification and Control: A Neural Network Approach*. Berlin: Springer-Verlag.

Ljung, L. (1987) *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall.

Ljung, L. and Södeström, T. (1983) *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press.

Marmarelis, P. and Marmarelis, V. (1978) *Analysis of Physiological Systems*. New York: Plenum Press.

Miller, W.T., Sutton, R.S., and Werbos, P.J. (eds) (1990) *Neural Networks for Control*. Cambridge, MA: MIT Press.

Nahi, N.E. (1976) *Estimation Theory and Applications*. New York: R.E. Krieger Publishing.

Nayfeh, A.H. and Mook, D.T. (1979) *Nonlinear Oscillations*. New York: John Wiley & Sons.

Nelles, O. (2001) *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin: Springer-Verlag.

Norton, J.P. (1986) *An Introduction to Identification*. New York: Academic Press.

Papoulis, A. (1965) *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill.

Pearson, R.K. (1999) *Discrete-Time Dynamic Models*. Oxford: Oxford University Press.

Peyton-Jones, J.C. and Billings, S.A. (1989) A recursive algorithm for computing the frequency response of a class of nonlinear difference equation models. *International Journal of Control, 50*, 1925–1940.

Pintelon, R. and Schoukens, J. (2001) *System Identification–a Frequency Domain Approach*. Oxford: Wiley-Blackwell.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, Vols *I and II*. New York: Academic Press.

Rugh, W.J. (1981) *Nonlinear System Theory—the Volterra/Wiener Approach*. Baltimore, MD: The Johns Hopkins University Press.

Schetzen, M. (1980) *The Volterra and Wiener Theories of Nonlinear System*. New York: John Wiley & Sons.

Södeström, T. and Stoica, P. (1989) *System Identification*. London: Prentice-Hall.

Swain, A.K., Billings, S.A., Stansby, P.K., and Baker, M. (1998) Accurate prediction of nonlinear wave forces: Part I. Fixed cylinder; Part II. Responding cylinder. *Journal of Mechanical Systems and Signal Processing, 12*, 449–485; 487–498.

Wei, H. and Billings, S.A. (2007) Feature subset selection and ranking for data dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 162–166.

Wellstead, P.E. and Zarrop, M.B. (1991) *Self Tuning Systems*. New York: John Wiley & Sons.

Wiener, N. (1958) *Nonlinear Problems in Random Theory*. Cambridge, MA: MIT Press.

Worden, K. and Tomlinson, G.R. (2001) *Nonlinearity in Structural Dynamics*. Bristol: Institute of Physics Publishing.

Worden, K., Stansby, P.K., Tomlinson, G.R., and Billings, S.A. (1994) Identification of nonlinear wave forces. *Journal of Fluids and Structures, 8*, 19–71.

Young, P.C. (1970) An instrumental variable method for real time identification of noisy processes. *Automatica, 6*, 271–287.

Young, P.C. (1984) *Recursive Estimation and Time-Series Analysis: An Introduction*. Berlin: Springer-Verlag.