Nonlinear System Identification

NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains

Stephen A Billings, University of Sheffield, UK

Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains describes a comprehensive framework for the identification and analysis of nonlinear dynamic systems in the time, frequency, and spatio-temporal domains. This book is written with an emphasis on making the algorithms accessible so that they can be applied and used in practice.

Includes coverage of:
- The NARMAX (nonlinear autoregressive moving average with exogenous inputs) model
- The orthogonal least squares algorithm that allows models to be built term by term where the error reduction ratio reveals the percentage contribution of each model term
- Statistical and qualitative model validation methods that can be applied to any model class
- Generalised frequency response functions which provide significant insight into nonlinear behaviours
- A completely new class of filters that can move, split, spread, and focus energy
- The response spectrum map and the study of sub harmonic and severely nonlinear systems
- Algorithms that can track rapid time variation in both linear and nonlinear systems
- The important class of spatio-temporal systems that evolve over both space and time
- Many case study examples from modelling space weather, through identification of a model of the visual processing system of fruit flies, to tracking causality in EGG data are all included to demonstrate how easily the methods can be applied in practice and to show the insight that the algorithms reveal even for complex systems

NARMAX algorithms provide a fundamentally different approach to nonlinear system identification and signal processing for nonlinear systems. NARMAX methods provide models that are transparent, which can easily be analysed, and which can be used to solve real problems. This book is intended for graduates, postgraduates and researchers in the sciences and engineering, and also for users from other fields who have collected data and who wish to identify models to help to understand the dynamics of their systems.
NONLINEAR SYSTEM IDENTIFICATION
All the world is a nonlinear system
He linearised to the right
He linearised to the left
Till nothing was right
And nothing was left
2.5 Block-Structured Models
   2.5.1 Parallel Cascade Models 32
   2.5.2 Feedback Block-Structured Models 32
2.6 NARMAX Models
   2.6.1 Polynomial NARMAX Model 35
   2.6.2 Rational NARMAX Model 37
   2.6.3 The Extended Model Set Representation 39
2.7 Generalised Additive Models 40
2.8 Neural Networks
   2.8.1 Multi-layer Networks 41
   2.8.2 Single-Layer Networks 42
2.9 Wavelet Models
   2.9.1 Dynamic Wavelet Models 46
2.10 State-Space Models 48
2.11 Extensions to the MIMO Case 49
2.12 Noise Modelling
   2.12.1 Noise-Free 50
   2.12.2 Additive Random Noise 50
   2.12.3 Additive Coloured Noise 50
   2.12.4 General Noise 51
2.13 Spatio-temporal Models 52
References 53

3 Model Structure Detection and Parameter Estimation 61
3.1 Introduction 61
3.2 The Orthogonal Least Squares Estimator and the Error Reduction Ratio 64
   3.2.1 Linear-in-the-Parameters Representation 64
   3.2.2 The Matrix Form of the Linear-in-the-Parameters Representation 65
   3.2.3 The Basic OLS Estimator 65
   3.2.4 The Matrix Formulation of the OLS Estimator 67
   3.2.5 The Error Reduction Ratio 68
   3.2.6 An Illustrative Example of the Basic OLS Estimator 69
3.3 The Forward Regression OLS Algorithm 70
   3.3.1 Forward Regression with OLS 72
   3.3.2 An Illustrative Example of Forward Regression with OLS 77
   3.3.3 The OLS Estimation Engine and Identification Procedure 78
3.4 Term and Variable Selection 79
3.5 OLS and Sum of Error Reduction Ratios 80
   3.5.1 Sum of Error Reduction Ratios 82
   3.5.2 The Variance of the s-Step-Ahead Prediction Error 82
   3.5.3 The Final Prediction Error 83
   3.5.4 The Variable Selection Algorithm 83
3.6 Noise Model Identification
   3.6.1 The Noise Model 84
   3.6.2 A Simulation Example with Noise Modelling 87
3.7 An Example of Variable and Term Selection for a Real Data Set 87
3.8 ERR is Not Affected by Noise 94
3.9 Common Structured Models to Accommodate Different Parameters 95
3.10 Model Parameters as a Function of Another Variable 98
3.10.1 System Internal and External Parameters 98
3.10.2 Parameter-Dependent Model Structure 98
3.10.3 Modelling Auxetic Foams – An Example of External Parameter-Dependent Model Identification 99
3.11 OLS and Model Reduction 100
3.12 Recursive Versions of OLS 102
References 102

4 Feature Selection and Ranking 105
4.1 Introduction 105
4.2 Feature Selection and Feature Extraction 106
4.3 Principal Components Analysis 107
4.4 A Forward Orthogonal Search Algorithm 108
4.4.1 The Basic Idea of the FOS-MOD Algorithm 108
4.4.2 Feature Detection and Ranking 109
4.4.3 Monitoring the Search Procedure 111
4.4.4 Illustrative Examples 112
4.5 A Basis Ranking Algorithm Based on PCA 113
4.5.1 Principal Component-Derived Multiple Regression 113
4.5.2 PCA-Based MFROLS Algorithms 114
4.5.3 An Illustrative Example 115
References 117

5 Model Validation 119
5.1 Introduction 119
5.2 Detection of Nonlinearity 121
5.3 Estimation and Test Data Sets 123
5.4 Model Predictions 124
5.4.1 One-Step-Ahead Prediction 124
5.4.2 Model Predicted Output 126
5.5 Statistical Validation 127
5.5.1 Correlation Tests for Input–Output Models 128
5.5.2 Correlation Tests for Time Series Models 132
5.5.3 Correlation Tests for MIMO Models 133
5.5.4 Output-Based Tests 134
5.6 Term Clustering 135
5.7 Qualitative Validation of Nonlinear Dynamic Models 137
5.7.1 Poincaré Sections 139
5.7.2 Bifurcation Diagrams 139
5.7.3 Cell Maps 140
5.7.4 Qualitative Validation in Nonlinear System Identification 140
References 145
## 6 The Identification and Analysis of Nonlinear Systems in the Frequency Domain

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>149</td>
</tr>
<tr>
<td>6.2 Generalised Frequency Response Functions</td>
<td>151</td>
</tr>
<tr>
<td>6.2.1 The Volterra Series Representation of Nonlinear Systems</td>
<td>153</td>
</tr>
<tr>
<td>6.2.2 Generalised Frequency Response Functions</td>
<td>156</td>
</tr>
<tr>
<td>6.2.3 The Relationship Between GFRFs and Output Response of Nonlinear Systems</td>
<td>157</td>
</tr>
<tr>
<td>6.2.4 Interpretation of the Composition of the Output Frequency Response of Nonlinear Systems</td>
<td>162</td>
</tr>
<tr>
<td>6.2.5 Estimation and Computation of GFRFs</td>
<td>165</td>
</tr>
<tr>
<td>6.2.6 The Analysis of Nonlinear Systems Using GFRFs</td>
<td>176</td>
</tr>
<tr>
<td>6.3 Output Frequencies of Nonlinear Systems</td>
<td>184</td>
</tr>
<tr>
<td>6.3.1 Output Frequencies of Nonlinear Systems under Multi-tone Inputs</td>
<td>185</td>
</tr>
<tr>
<td>6.3.2 Output Frequencies of Nonlinear Systems for General Inputs</td>
<td>187</td>
</tr>
<tr>
<td>6.4 Nonlinear Output Frequency Response Functions</td>
<td>191</td>
</tr>
<tr>
<td>6.4.1 Definition and Properties of NOFRFs</td>
<td>192</td>
</tr>
<tr>
<td>6.4.2 Evaluation of NOFRFs</td>
<td>195</td>
</tr>
<tr>
<td>6.4.3 Damage Detection Using NARMAX Modelling and NOFRF-Based Analysis</td>
<td>196</td>
</tr>
<tr>
<td>6.5 Output Frequency Response Function of Nonlinear Systems</td>
<td>202</td>
</tr>
<tr>
<td>6.5.1 Definition of the OFRF</td>
<td>203</td>
</tr>
<tr>
<td>6.5.2 Determination of the OFRF</td>
<td>203</td>
</tr>
<tr>
<td>6.5.3 Application of the OFRF to Analysis of Nonlinear Damping for Vibration Control</td>
<td>207</td>
</tr>
</tbody>
</table>

References 213

## 7 Design of Nonlinear Systems in the Frequency Domain – Energy Transfer Filters and Nonlinear Damping

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>217</td>
</tr>
<tr>
<td>7.2 Energy Transfer Filters</td>
<td>218</td>
</tr>
<tr>
<td>7.2.1 The Time and Frequency Domain Representation of the NARX Model with Input Nonlinearity</td>
<td>220</td>
</tr>
<tr>
<td>7.2.2 Energy Transfer Filter Designs</td>
<td>222</td>
</tr>
<tr>
<td>7.3 Energy Focus Filters</td>
<td>240</td>
</tr>
<tr>
<td>7.3.1 Output Frequencies of Nonlinear Systems with Input Signal Energy Located in Two Separate Frequency Intervals</td>
<td>241</td>
</tr>
<tr>
<td>7.3.2 The Energy Focus Filter Design Procedure and an Example</td>
<td>245</td>
</tr>
<tr>
<td>7.4 OFRF-Based Approach for the Design of Nonlinear Systems in the Frequency Domain</td>
<td>249</td>
</tr>
<tr>
<td>7.4.1 OFRF-Based Design of Nonlinear Systems in the Frequency Domain</td>
<td>249</td>
</tr>
<tr>
<td>7.4.2 Design of Nonlinear Damping in the Frequency Domain for Vibration Isolation: An Experimental Study</td>
<td>251</td>
</tr>
</tbody>
</table>

References 259
10.2 The Kernel Invariance Method
10.2.1 Definitions
10.2.2 Reconstructing the Linear Model Terms
10.2.3 Reconstructing the Quadratic Model Terms
10.2.4 Model Structure Determination

10.3 Using the GFRFs to Reconstruct Nonlinear Integro-differential Equation Models Without Differentiation
10.3.1 Introduction
10.3.2 Reconstructing the Linear Model Terms
10.3.3 Reconstructing the Quadratic Model Terms
10.3.4 Reconstructing the Higher-Order Model Terms
10.3.5 A Real Application

References

11 Time-Varying and Nonlinear System Identification
11.1 Introduction
11.2 Adaptive Parameter Estimation Algorithms
11.2.1 The Kalman Filter Algorithm
11.2.2 The RLS and LMS Algorithms
11.2.3 Some Practical Considerations for the KF, RLS, and LMS Algorithms

11.3 Tracking Rapid Parameter Variations Using Wavelets
11.3.1 A General Form of TV-ARX Models Using Wavelets
11.3.2 A Multi-wavelet Approach for Time-Varying Parameter Estimation

11.4 Time-Dependent Spectral Characterisation
11.4.1 The Definition of a Time-Dependent Spectral Function

11.5 Nonlinear Time-Varying Model Estimation

11.6 Mapping and Tracking in the Frequency Domain
11.6.1 Time-Varying Frequency Response Functions
11.6.2 First and Second-Order TV-GFRFs

11.7 A Sliding Window Approach

References

12 Identification of Cellular Automata and N-State Models of Spatio-temporal Systems
12.1 Introduction
12.2 Cellular Automata
12.2.1 History of Cellular Automata
12.2.2 Discrete Lattice
12.2.3 Neighbourhood
12.2.4 Transition Rules
12.2.5 Simulation Examples of Cellular Automata

12.3 Identification of Cellular Automata
12.3.1 Introduction and Review
12.3.2 Polynomial Representation
12.3.3 Neighbourhood Detection and Rule Identification
14.8 Real-Time Diffuse Optical Tomography Using RBF Reduced-Order Models
of the Propagation of Light for Monitoring Brain Haemodynamics 514
14.8.1 Diffuse Optical Imaging 515
14.8.2 In-vivo Real-Time 3-D Brain Imaging Using Reduced-Order Forward Models 517
14.9 Identification of Hysteresis Effects in Metal Rubber Damping Devices 522
14.9.1 Dynamic Modelling of Metal Rubber Damping Devices 523
14.9.2 Model Identification of a Metal Rubber Specimen 526
14.10 Identification of the Belousov–Zhabotinsky Reaction 528
14.10.1 Data Acquisition 529
14.10.2 Model Identification 530
14.11 Dynamic Modelling of Synthetic Bioparts 534
14.11.1 The Biopart and the Experiments 535
14.11.2 NARMAX Model of the Synthetic Biopart 536
14.12 Forecasting High Tides in the Venice Lagoon 539
14.12.1 Time Series Forecasting Problem 540
14.12.2 Water-Level Modelling and High-Tide Forecasting 541
References 543

Index 549
System identification is a method of identifying or measuring the dynamic model of a system from measurements of the system inputs and outputs. System identification was developed as part of systems and control theory and has now become a toolbox of algorithms and methods that can be applied to a very wide range of real systems and processes. The applications of system identification include any system where the inputs and outputs can be measured. Applications therefore include industrial processes, control systems, economic data and financial systems, biology and the life sciences, medicine, social systems, and many more.

System identification has become an important topic across many subject domains over the last few decades. Initially, the focus was on linear system identification but this has been changing with more of an emphasis on nonlinear systems over recent years. There are several excellent textbooks on linear system identification, time series, spectral analysis methods and algorithms, and hence there is no need to repeat these results here. Rather, the focus of this book is on the identification of nonlinear dynamic systems using what have become known as NARMAX methods. NARMAX, which stands for a nonlinear autoregressive moving average model with exogenous inputs, was initially introduced as the name of a model but then developed into a framework for the identification of nonlinear systems. There are other methods of nonlinear system identification, and many of these are also discussed within the book. But NARMAX methods are based on the goal of determining or identifying the rule or law that describes the behaviour of the underlying system, and this means the focus is on determining the form of the model, what terms should be included in the model, or the structure of the model. The focus is therefore not on gross approximation but on identifying models that are as simple as possible, models that can be written down and related to the underlying system, and which can be used to tease apart and understand complex nonlinear dynamic effects in the wide range of systems that system identification can be applied to.

At the core of NARMAX methods is the ability to build models by finding the most important term and adding this to the model, then finding and adding the next most important term, and so on so that the model is built up in a simple and intuitive way. This mimics the way traditional analytical modelling is done, by finding the most important model terms and then
building the model up step by step until a desired accuracy is achieved. The difference with NARMAX methods is that this process is accomplished using measured data in the presence of possible nonlinear and highly coloured noise. The concepts behind this process are simple, intuitive, and easy to use.

There is extensive research literature in the form of published papers on many aspects of nonlinear system identification, including NARMAX methods. The aim in this book is not to reproduce all the many variants of the algorithms that exist, but rather to focus on presenting some of the best algorithms in a clear way. All the detailed nuances and variants of the algorithms will be cited within the book, so that anyone with more theoretical interests can follow up these ideas. But the aim of this book is to focus on the core methods, to try to describe them using the simplest possible terminology, and to clearly describe how to use them in real applications. This will inevitably involve mathematical descriptions and algorithmic details, but the aim is to keep the mathematics as simple as possible. The core aim therefore is to write a book that readers from a range of disciplines can use to understand how to fit models of dynamic nonlinear systems.

The book is an attempt to fill a void in the existing literature. Currently, there are several books on neural networks, and all the variants of these, and on the identification of simple block-structured nonlinear systems. These are important topics, but they address essentially different problems than the main aim of this book. Neural networks are excellent for fitting models for prediction purposes, but they do not produce transparent models, models that can be written down, and which can be analysed in time and frequency. Block-structured systems are a special class of nonlinear systems which are all based on the assumption that the system under study is a member of this simple class.

The main aim of this book is to describe a comprehensive set of algorithms for the identification and analysis of nonlinear systems in the time, frequency, and spatio-temporal domains. While almost every other textbook on nonlinear system identification is focused on time domain methods, we want to address the total oversight in the literature and include frequency and spatio-temporal methods which can provide significant insights into complex system behaviours. These are natural extensions of NARMAX identification methods and offer new directions in nonlinear system identification with many applications.

The readership will include graduates, postgraduates, and researchers in the sciences and engineering, but also users from other research fields who have collected data and who wish to identify models to help understand the dynamics of their systems. While there are examples throughout the book, the last chapter contains many case studies. These are used to illustrate how the methods described in the book can be applied to a wide range of problems from modelling the visual system of fruit flies, to detecting causality in EGG signals, modelling the variations in ice flow, and modelling space weather. These examples are included to demonstrate that the methods in this book do work, that models can quite easily be identified in an intuitive and straightforward way, and used to understand and gain new insights into what appear to be complex effects.

The book starts in Chapter 1 where the focus of the book, the context in which the methods were developed, and the reason for the approaches taken are described in detail. Chapter 2 introduces the different classes of dynamic models. Chapter 3 describes model structure detection and parameter estimation based on the orthogonal least squares (OLS) algorithm and the error reduction ratio. Chapter 4 shows how the methods of Chapter 3 can be adapted for feature and basis function selection. Chapter 5 discusses model validation. Chapter 6 introduces important concepts for the frequency domain analysis of nonlinear systems, and Chapter 7 builds on these results to describe a new class of filters that can be designed to move energy to desired frequency
locations, and the design of nonlinear damping devices. Chapter 8 describes how neural networks, including radial basis function and wavelet networks, can be used in system identification. Chapter 9 discusses the identification and analysis of severely nonlinear systems. Chapter 10 is focused on the identification of continuous-time nonlinear models. Chapter 11 shows how very rapid time variation in nonlinear models can be identified and tracked in both time and frequency. Chapter 12 describes spatio-temporal systems with finite states, including cellular automata models and $n$-state models, and the identification of these. Chapter 13 describes the spatio-temporal class of systems that have a continuous state and introduces system identification, analysis, and frequency response methods for this important class of systems. Chapter 14 includes a very wide range of case studies relating to many important problems.

A graduate course of 20–30 hours could be built using sections from the book. Such a course might include the core models from Chapter 2, the basic and forward regression orthogonal least squares algorithm and the error reduction ratio test from Chapter 3, brief details of feature extraction from Chapter 4, the simple correlation model validity tests for nonlinear systems from Chapter 5, the introduction of generalised frequency response functions and the estimation and interpretation of these using the simple probing methods from Chapter 6, radial basis function neural network training and input node selection using orthogonal least squares concepts from Chapter 8, wavelet models and the response spectrum map from Chapter 9, an introduction to spatio-temporal systems based on cellular automata and coupled map lattice models from Chapters 12 and 13, and finally some case study examples from Chapter 14.

I would like to acknowledge all those who have supported me over many years, those that I have worked with and learnt from, and those that have helped to write each chapter in this book. This book could not have been written without considerable help from colleagues. I would like to acknowledge this help by thanking Hualiang Wei who contributed Chapters 2, 3, 4, 5, 8, and 11; Zi Qiang Lang for Chapters 6 and 7; Liangmin Li for Chapters 9 and 10; Yifan Zhao for Chapter 12; Lingzhong Guo for Chapter 13; and Otar Akanyeti, Misha Balikhin, Richard Boynton, Yifan Zhao, Hualiang Wei, Uwe Friedrich, Danial Coca, Ernesto Vidal Rosas, Bin Zhang, Krish Krishnanathan, and Visakan Kadirkamanathan for help with the case studies.

Over many years I have supervised over 50 PhD students and worked with a similar number of research assistants. I have also been supported, challenged, and inspired by many academic colleagues and friends, both within my own discipline and in other research fields. There are too many to name but they all made important contributions which I would like to acknowledge. Although I can find no records now, my recollection is that Cristian Georgescu supplied the poem about nonlinearity in a personal communication when he applied to study for a PhD with me but unfortunately could not take up this position.

Much of the work in this book has been achieved with support from the research councils and other funding bodies. I gratefully acknowledge this support from the Engineering and Physical Sciences Research Council (EPSRC), The European Research Council (ERC), the Biotechnology and Biological Research Council (BBSRC), the Natural and Environment Research Council (NERC), and the Leverhume Trust.

I would like to especially thank all my family, Professor Harry Nicholson, Duncan Kitchen, Alan and Joyce Bellinger, the medics and nurses, and all those who gave unremitting support during a life-threatening illness. Finally, I would like to thank all my family for their support during my early education and throughout my career, I am especially grateful for this constant support.

This book is dedicated to my late father George Billings, who taught me without really teaching.
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
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 Ohm’s 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 Eykhoff, 1971; Graupe, 1972; Eykhoff, 1974; Nahi, 1976; Godwin 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 correlated 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 prediction 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 density 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 nonlinear systems, which can be tightly defined, but which are limited. Historically, system identification 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

\[
y(k) = h_0 + \sum_{m_1=1}^{M} h_1(m_1) u(k - m_1) + \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} h_2(m_1, m_2) u(k - m_1) u(k - m_2) \\
+ \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} \sum_{m_3=1}^{M} h_3(m_1, m_2, m_3) u(k - m_1) u(k - m_2) u(k - m_3) + \cdots
\]  

(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 \( \ell \)th-order Volterra kernel, or \( \ell \)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, \( 30 \times 30 \) 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 func-
tions, which could determine which of the block-structured models was appropriate to repre-
sent 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 iden-
tification 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

\[ y(k) = F[y(k-1), y(k-2), \ldots, y(k-n_y),
                    u(k-d), u(k-d-1), \ldots, u(k-d-n_u),
                    e(k-1), e(k-2), \ldots, e(k-n_e)] + e(k) \] (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

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

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

\[ y(k) + a_1 y(k-1) + \cdots + a_{na} y(k-na) = b_1 u(k-1) + \cdots + b_{mb} u(k-nb) \] (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 rep- sons 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 sys- tem. 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 non- linear, 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 mod- els 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 underly- ing 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