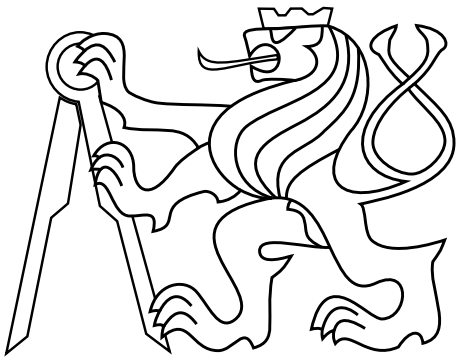
DISCRETE WAVELET TRANSFORM IN LINEAR SYSTEM IDENTIFICATION

Zdeněk Váňa



Doctoral Thesis

Department of Control Engineering
Faculty of Electrical Engineering
Czech Technical University in Prague

Prague, February 2014

PhD Program:
Electrical Engineering and Information Technology

Branch of Study:
Control Engineering and Robotics

Supervisor:
Prof. Ing. Michael Šebek, DrSc.
Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague

Supervisor-specialist:
Doc. Ing. Lukáš Ferkl, Ph.D.
Department of Control Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague

Education is the best provision for the journey to old age.

— Aristotle

The most incomprehensible thing about the world is that it is comprehensible.

— Albert Einstein

This work is dedicated to my family.

Declaration

This doctoral thesis is submitted in partial fulfillment of the requirements for the degree of doctor (Ph.D.). The work submitted in this dissertation is the result of my own investigation, except where otherwise stated.

I declare that I worked out this thesis independently and I quoted all used sources of information in accord with Methodical instructions about ethical principles for writing academic thesis. Moreover I declare that it has not already been accepted for any degree and is also not being concurrently submitted for any other degree.

Prague, February 2014

Zdeněk Váňa

Acknowledgements

Foremost, I would like to sincerely thank to my supervisors Prof. Ing. Michael Šebek, DrSc. and Doc. Ing. Lukáš Ferkl, Ph.D. for their excellent leadership and support, for providing me with the unique conditions for research and for sharing their ideas with me.

A special thanks belong to Prof. Heinz A. Preisig from Norwegian technical university NTNU for giving me advices when it was mostly needed and for creating conditions for perfect co-operation.

Finally, my immense gratitude goes to my parents as well as to my girlfriend Martina Babjáková for their endless support and encouragement.

Then, the heartfelt thanks belong to my colleagues Samuel Prívara and Jiří Cigler for their time, advices, critical thinking and for sharing many common-life problems. I must also highlight a lot of free time we have spent together and many experiences we have gained together.

I am grateful to my colleagues and friends Ondřej Nývlt, Eva Žáčeková, Matej Pčolka, Jan Šulc, David Charvát, Jan Široký from Energocentrum PLUS, s.r.o. and Josef and Dana Komárkovi from Technofiber, s.r.o. for a friendly atmosphere and working conditions I could work in. You indeed were the best colleagues I could ever imagine.

I am sure that there are many people which are not explicitly listed here. Nevertheless, I am very grateful for everything they have done for me and they stay in my thoughts.

At last, but not the least, this thesis would not have been the same without the financial support from the state budget of the Czech Republic, through the Ministry of industry and commerce, in the scope of grant No. 5.1 SPK01/024 – "ENERGOKLASTR", in the scope of grant No. FR-TI1/517, "Control Systems for Optimization of Energy Consumption in Low-Energy and Passive Houses', in the scope of grant No. 2A-1TP1/084, "Integration of building systems, research and application of intelligent algorithms with

influence on energy consumption of buildings and living houses", through the Grant Agency of the Czech Republic (GACR) in the scope of grant No. P103/12/1187 and finally through grant CZ.1.05/2.1.00/03.0091 – "University Centre for Energy Efficient Buildings (UCEEB)".

Abstract

A system identification plays a central role in any activity associated with the process control. As the usual purpose of the model is its consequent use for control, the system identification should be focused mainly on those dynamics, which are important from the control point of view. Moreover, with the control being done on different levels, different models can be required for the same plant, each for a different range of dynamics.

A wavelet transform is quite young mathematical tool for signal analysis. It analyses a signal in both time and frequency domain at once and is being performed via wavelet functions. Due to an intrinsic localisation of the wavelet functions in time and frequency, methods exploiting wavelets may, to some extent, represent a natural way of selection of an important information hidden in a data. Consequently, a system identification adopting wavelets introduces a possibility to obtain a model describing the plant from important dynamics point of view only, hence a model suitable for control. It is also very convenient for identification of the dominant modes of the system, e.g. when identifying singularly perturbed systems.

Some early works in system identification showed that a linear model with a suitable structure and an appropriate order can sufficiently approximate any dynamics, even non-linear. The mostly used methods intended for an identification of linear models however do not provide the user with some simple approach to select particular range of dynamics to be identified, but with the only direct computation of a filter and subsequent filtering of data.

The thesis presents several approaches to system identification in which wavelet transform is employed for both single and multivariable system identification enabling selection of the particular frequency range of interest. Moreover, the thesis treats the possibility of applying the wavelets within both continuous-time and discrete-time system identification. Next, as there are several families of wavelet functions, each possessing different properties, the incorporation of wavelet transform into a system identification is treated in a general concept. However, besides others, there are two divisions of wavelet functions, each in accordance to one particular property. The first property is a compact support in time domain and still good selectivity in

frequency domain. These wavelets are utilized with advantage of accurate implementation. The second property is a mutual orthogonality of wavelets, what leads to exploiting the data without any loss of information. The thesis also provides several points of view on wavelets, what enables the reader to understand both the theory of wavelets and system identification theory more deeply. A lot of connections of both theories are very intuitive, hence are simply applicable in different fields.

Acronyms

ARMAX Auto-Regressive Moving Average with eXternal input

ARX Auto-Regressive with eXternal input

BJ Box Jenkins

CTU Czech Technical University

CWT Continuous Wavelet Transform

DFT Discrete Fourier Transform

DWT Discrete Wavelet Transform

ETFE Empirical Transfer Function Estimate

FIR Finite Impulse Response

FT Fourier Transform

LTI Linear Time-Invariant

LTV Linear Time-Varying

MIMO Multiple-Input Multiple-Output

MISO Multiple-Input Single-Output

MRA MultiResolution Analysis

OE Output Error

OLS Ordinary Least Squares

PEM Prediction Error Methods

PMFD Polynomial Matrix Fractional Description

SID System IDentification

SISO Single-Input Single-Output

STFT Short-Time Fourier Transform

WT Wavelet Transform

Mathematical notations

The list of mathematical notation used within the thesis follows. Note that the notation holds unless otherwise stated.

$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integer numbers
$\mathbb{R}$	Set of real numbers
$\ell^2(\mathbb{Z}_N)$	Space of complex sequences of N-dimensional vectors
$\ell^2(\mathbb{Z})$	Space of complex square-summable sequences
$L^2(\mathbb{R})$	Space of complex square-integrable functions
t, k	Time - either continuous or discrete
ω	Frequency - either continuous or discrete
φ	General notation of basic father wavelet
ψ	General notation of basic mother wavelet
L	Length of wavelet filter (either father or mother)
N	Length of data to be analysed
p	Maximum level of wavelet analysis
T	Wavelet matrix
W	Weighting matrix
$\lfloor z \rfloor$	Integer part of number z
$\oplus$	Direct sum
$\otimes$	Kronecker product
$\hat{\bullet}$	When $\bullet$ stands for a complex function, $\hat{\bullet}$ denotes its Fourier transform. Otherwise it denotes the estimate of the true value of $\bullet$.
n_a	Order of the polynomial a
n_b	Order of the polynomial b
m	Number of system's inputs
n	Number of system's states or order of the system
o	Number of system's outputs
θ	Vector of unknown model's parameters
A, B, C, D	Matrices of system's or model's state-space description
G, H	Transfer functions of deterministic and stochastic part, respectively

Contents

List of Figures

List of Tables

1 Introduction

Mathematical modeling of systems and processes is an essential tool used in science and engineering from its very beginning. Newton's laws are an example for a model. They describe the motion of bodies such as the various bodies in the solar system: the Sun, the planets and their moons. Models are used extensively in all branches of science and engineering, which raised to the term model-based engineering. The fitting of the model to the actual physical system takes a central position, which systems engineers refer to as system identification, and it gained a lot of attention until quite recently in control engineering. The development of model-based control is originated in 1960's and it is one of the mile-stones in control. Before that time, most of control design strategies had been based either on heuristic methods (Ziegler-Nichols method for tuning of PID controller parameters) or on system properties obtained via simulation or measurement (impulse response, Bode graph). Since 1960's, a huge development of control techniques employing the model of the controlled process has been recorded and hand in hand with this development, system identification became a more important part of control theory. Nowadays, system identification is an indispensable part of modern control theory and covers a wide range of problems and approaches to their solution. Moreover, the progress in system identification is still under pressure, since the growing requirements on control naturally reflect into the growing demands on the model. Details about the history of system identification can be found in Gevers [2006].

Model is an image of the real process that represents its properties essential to the application and that behaves within a given context similar to the real process. There is, however, a plenty of choices for the modeling itself (a description of the model, a model structure or a model order), which affect the resulting model and its properties. Therefore, there can be several models of the same process, each describing the process in a different way, hence the model of the process is by no means unique. As the model is used for a subsequent control, it is obvious that it should reflect only those properties of the system, which are important from the control point of view. In addition, the model should also be of a sufficient quality to be suitable for a control.

The process of system identification can be considered as follows. i) To acquaint with the application and to select its inputs and outputs, which are necessary for control. ii) To select a model structure, possibly in accordance to the control strategy. iii) Based on the model structure, to choose the algorithm for a system identification and to specify model parameters to be estimated. iv) To select the criterion to measure the quality of the model. v) To estimate the parameters of the model. vi) And finally, to validate the model. However, in case of an insufficient model quality, the procedure is repeated until a good model is obtained.

The model parameters estimate is computed from measured input and output data and since these measurements record both wanted and unwanted process behavior, it is therefore very suitable to analyse the data and extract the appropriate portion of information from them. One of possibilities of data analysis is the Wavelet Transform (WT), the transform possessing several advantageous features and bringing new ideas into the concept of system identification. Further on, a brief story about the origin of the WT is written as well as the motivation why the WT utilization should be investigated.

1.1 Origins of wavelet analysis

One of the most used signal analysis had been introduced by French mathematician and physicist Jean Baptiste Joseph Fourier (∗1768; † 1830) and, in his honor, has been called Fourier analysis (or Fourier Transform (FT)). Fourier analysis reflects the time domain into the frequency domain and results in frequency spectra of the time-varying signal. The natural disadvantage of this transform is the fact, that one domain disables detection of important phenomena in the second domain and vice versa - one is not possible to detect important time instants from the frequency spectra as well as to specify main frequencies (or harmonic functions) from the time domain. Many years after Fourier, Hungarian electrical engineer and physicist Denis Gabor (∗1900; † 1979) comes up with Short-Time Fourier Transform (STFT), where the FT of the signal is applied within the given time-window only. STFT is thus trade-off between the time and the frequency signal description, however, its limiting factor is type and mainly fixed size of time-window. As a next logical step, STFT with time-varying window has been inspected what laid the fundamental idea of wavelet analysis. WT analyses the signal from both time and frequency points of view. In principle, WT is similar to STFT. The main difference is that the time-window is not fixed, but scaled across the levels of WT.

In late 1970's, French geophysicist Jean Morlet (∗1931; † 2007) worked on an analysis of the reflected signal containing both high and low frequency content on short and long time spans, respectively. During this work he noticed that keeping the time-window of STFT fixed is wrong approach. Therefore he came up with the idea of changing the time-window while keeping its frequency content. What Morlet did is that, while in STFT the signal is multiplied by particular time-window and then anal-

ysed by all harmonic functions one by one, he multiplied particular harmonic function by time-window (what actually was a Gaussian envelope) and the resulting function is used for the analysis. Because of the shape of this new function Morlet called it "wavelet" and this term is used till today. On Morlet's honor, harmonic function multiplied by Gaussian window is nowadays called Morlet's wavelet.

In 1980, Morlet met the Croatian-French physicist Alexander Grossmann (∗1930) and together rigorously formalized basics of wavelet analysis what included stating the wavelets by its properties. Moreover, Grossmann helped the Morlet with development of inverse wavelet transform. Next two great contributors to the field of wavelets were French mathematician and scientist Yves Meyer (∗1939) and, in that time student, Stéphane Mallat, who developed the idea of multiresolution analysis. That was a big step in a wavelet theory, since the scaling of wavelet functions was introduced for the first time. It consequently enabled to construct a general wavelet theory independently on the wavelet functions. During derivation of the multiresolution analysis, an interesting fact was found out, namely that wavelet history goes back to 1909, when Hungarian mathematician Alfred Haar (∗1885; † 1933) - a student of German mathematician David Hilbert - created a simple orthogonal system of functions that satisfies both the wavelet properties and conditions for multiresolution analysis, thus became the first and the simplest system of wavelets at the same time. The last significant contributor and wavelet researcher who should not be forgotten is Belgian mathematician and physicist Ingrid Daubechies (∗1954), who in late 1980's used the multiresolution analysis to create new system of wavelet functions. These wavelets are called Daubechies wavelet and have many important properties for which became the most used wavelets after the Haar wavelet. For further details on history of WT see e.g. Polikar [1999].

1.2 Motivation for wavelet transform

Wavelet transform, from its pure essence, brought a lot of possibilities into the science. As every new mathematical tool, it introduced a new way of describing of some parts of the nature and enabled to mathematically formulate some scientific problems. As was already mentioned, the main advantage of WT is its possibility to analyse the signal in both time and frequency domains. This is, actually, not only the advantage, but simultaneously quite important feature and it should be understood correctly. Wavelet analysis arose from STFT by generalization of the time-window onto wavelets, which is well localised both in time and frequency. It consequently means that instead of the signal being described as a function of either time or frequency, it can be viewed in both time and frequency simultaneously. Analysis in time proceeds by shifting the wavelet along the time axis and analysis in frequency proceeds by scaling the wavelet.

This dual time-frequency approach to signal analysis is, of course, counterbalanced by more complicated mathematical background, however, just thanks to this duality WT found a way to plenty of real applications as one of the most convenient mathe-

matical instrument. As WT serves mainly as a tool for signal analysis, typical ways of its utilization belong into the signal processing, namely i) detection of signal discontinuities, ii) trend detection, iii) detection of self-similarities, iv) particular frequency detection, v) signal suppression, vi) signal denoising and vii) data compression. Although these above-mentioned ways are widely known and used, applications of WT are not limited to them at all. The analysis itself is not the only new product of wavelet research, but the wavelet functions as well. There are many kinds of wavelet functions each of them possess several desired and useful properties. Thus the signal analysis does not have to be the only purpose of utilizing of wavelets. Additionally, as time went on, it showed up that one can look at the wavelets from several different points of view. While classical approaches to wavelets are via functional analysis or vector spaces theory, other concepts were proposed, e.g. through theory of frames in vector spaces, set theory or theory of finite elements.

As was mentioned above, wavelets with their characteristic properties have number of possible applications in many diverse fields, usually those closely related to the scientific research and development. Some obvious applications are biological signals analysis (EEG, EKG), analysis of seismic activity and prediction of earthquakes, analysis of sounds, multidimensional signal analysis together with data compression often applied to image processing as specific representative, analysis and attempts to predict the behavior of stock market, financial data analysis, etc. This all denote the capabilities of WT and wavelets themselves and predict them great importance in the future. However, despite the apparent strength of WT, it is still not used as much and as frequently as it can be and as it deserves.

1.3 Organization of the Thesis

This thesis is further structured as follows. Chapter 2 states the aims of the thesis. In Chapter 3, a comprehensive survey of the state of the art of using wavelets within the system identification is presented. Chapter 4 and Chapter 5 introduce briefly both wavelet theory and system identification theory, respectively. Three main parts of the thesis follow. The first of them, Chapter 6, treats the utilization of wavelet transform within a single LTI system identification, discuses different setups of the algorithm and ways of its adapting to the real application. Moreover, asymptotic properties of the proposed algorithm are discussed as well as some possible extending approaches. Next, Chapter 7 extends the Chapter 6 for multivariable systems. Three methods for the multivariable system identification are given together with their adaptation to the wavelet transform incorporation. Yet another method for utilizing the wavelet transform for system identification is presented in Chapter 8. Despite the fact that there were some attempts in the field of modulating functions and wavelet theory (see the state of the art in Chapter 3), to the author's best knowledge no general interconnection has not been done, therefore the Chapter 8 dealing with this topic was the must for this thesis. Chapter 9 concludes the thesis.

Each of main chapters (Chapter 6, Chapter 7 and Chapter 8) begin with a brief introducing of the system identification method and then its adaptation to the wavelet transform follows. Moreover, in Chapter 6 and Chapter 7, a functionality of the proposed algorithms is shown on a suitable case studies. In Chapter 8, the case study is missing since its functionality is in principle already shown within the case studies in Chapter 6 and Chapter 7 as the algorithms share the basic idea. Chapters 6 and 7 were partially published within the author's publications.

2 Aims of the Thesis

The thesis is entitled Discrete Wavelet Transform in Linear System Identification and since it is a very general topic, it had to be studied before the work on thesis started. The theory of linear system identification became very important when the model-based control theory has arisen and is now widely used in practise. Next, a search for a good models that are suitable for control applications yields the fact that both academicians and engineers develop or adapt the identification methods for their specific application. An alternative to developing new theories is to combine seemingly unrelated theories thereby adding new components to the identification procedures and adding new views on the involved theories.

Therefore, the main theme of the thesis lies in the interconnection of wavelet theory and theory of linear system identification. The goals are split into the following subjects:

1. To perform a comprehensive survey of the methods of exploiting the wavelet transform for system identification.

2. To find and describe a suitable way of incorporation of wavelet transform into the problem of general single-input single-output linear system identification. Analyse the method and demonstrate it on a suitable example.

3. To extend the method to multivariable systems. Analyse the method and demonstrate it on a suitable example.

4. To investigate and find the utilization of wavelet transform within the continuous-time linear system identification. The discussion on implementation issues must be included.

3 State of the art

Until recently, wavelet theory has not been extensively applied to many theoretical and practical problems belonging to the system identification so much. However, there are several papers which deal with application of wavelets in this field. Almost all of that applications exploit the superior properties of wavelet analysis or are at least based on some specific property of the particularly used wavelet function - both cases have already been discussed in Chapter 1. Moreover, absolute majority of that applications deal with a specific wavelet function only and do not consider general wavelet function.

In this chapter, we will go through the most used, important and meaningful wavelets utilisations in system identification in more details. And as the wavelets enters distinct areas of research, we will sort them in order to the object of the research rather than along the time axis.

3.1 Wavelet transform in linear system identification

3.1.1 2^{nd} order systems identification

Starting with simple Linear Time-Invariant (LTI) systems, one of the first attempts was to estimate oscillatory properties of a system like natural frequency, damping and stiffness. Ruzzene et al. [1997] proposed a wavelet based estimation method for the system of several interconnected oscillatory 2^{nd} order systems. For the estimation, the Morlet's wavelets were used because of their advantage of simple mathematical formulation.

Next, Boltezar and Slavic [2004] being inspired by the previous paper solved a similar problem, but with the help of parametrized Gaussian windows instead of Morlet's wavelet. The idea of wavelet edges showing the importance of particular wavelet coefficients was utilized, however, usual method based on ordinary edge-effect (Staszewski [1998]) was shown to be unsuitable. The authors thus came up with three new approaches to the edge-effect, all improving the proportionality between the wavelet coefficients and the analysed signal.

Further works dealing with the estimating of the 2^{nd} order system parameters are

e.g. Erlicher and Argoul [2007]; Huang and Su [2007]; Joo [2012]; Kijewski and Kareem [2003]; Kougioumtzoglou and Spanos [2013]. It is worth to note that many of them use the Continuous Wavelet Transform (CWT).

3.1.2 Wavelet analysis of system relevant signals

Another method of employing wavelets for system identification is to apply the WT to the system relevant signals. Such a general characterization is used intentionally, since there are several different utilisations of wavelet analysis for system inputs, outputs or states across the literature. Some of them even misuse the wavelet theory or its notation, so they are not either correct from the rigorous mathematical point of view on WT or well-advised from an engineering point of view, read e.g. Section 3.1.4. Nevertheless, leaving the mathematical accuracy, those papers bear with no doubts interesting ideas.

Luk and Damper [2006] exploited one of the fundamental properties of wavelets - their mutual orthogonality - to design a suitable system input. It yields the inverse WT of the system's impulse response. The mutual orthogonality of wavelet functions plays an important role also in the Serban [2007], where the authors took advantage of this property to decompose the input and output of the system into more signals, each lying in specific frequency range. Several models thus could be identified each describing a distinct part of the overall dynamics. Morlet's wavelets were used.

The ability of removing noise has also found attention: Wang et al. [2010] denoised system input and output to suppress the high frequency content of data to improve the accuracy of parameter estimation.

At last, a general framework of applying the WT on system inputs and outputs should be mentioned, see e.g. Carrier and Stephanopoulos [1998]; Erlicher and Argoul [2007]; Kinoshita and Ohta [2010]; Lardies et al. [2004]; Mukhopadhyay and Tiwari [2010]; Shan and Burl [2011]; Shiguo et al. [2004]; Xu et al. [2012]. This framework does not, actually, contain any specific procedure at all, but represents the natural way of utilization of WT. It is used across different applications of WT no matter the problem specification or complexity. A quite surprising fact is that not only are advantages of wavelets the reason for using the WT for improving either the identification method or the resulted model, but also endeavour or enthusiasm are other frequently occurred reasons for using of WT. In general, the later reasons are usual for investigating a new way or trying to look at the problem from a different point of view. Indeed, those reasons are not explicitly mentioned in any paper, however it is important human nature which forces us to do so even without any well-founded reason.

3.1.3 Wavelet as modulating functions

When studying the WT applications for linear systems identification we should not forget to pay attention to the relationship of wavelets and modulating functions. The basic principle of applying modulating function for system identification is well established topic and firstly was suggested by Shinbrot [1954], however for more details the reader

can look for some recent book, e.g. Rao and Unbehauen [2006]. Modulating functions satisfy 3 specific properties and there are some wavelet functions which comply them as well.

A great summarizing study about the usage of the modulating functions on the system identification has been performed in triplet Preisig and Rippin [1993a,b,c], where mostly Maletinsky's splines were used as modulating functions. Briefly, the first part introduces a general modulating function method and provides a history survey, the second part shows an algebraic formulation of Maletinsky's splines and extends the method to the arbitrarily overlapping modulating functions. The third part then apply the extended method on a batch process. General spline approach to wavelets is treated in Chui and Wang [1991, 1992].

A few years later, one of the first attempts to employ wavelet function as a modulating function was doubtless Kosanovich et al. [1995]. The Poisson wavelet was used here, which is derived from the Poisson probability density function being a kernel of Poisson transform. Note that Poisson wavelet is neither compactly supported (due to the exponential in its expression) nor orthogonal, thus can not be considered as a real wavelet or modulating function, although the previous related work of Kosanovich had shown that Poisson wavelet complies any of wavelet properties. However, when approximating the exponential, one obtains a function which satisfies desired properties. Consequential work Ramarathnam and Tangirala [2009] analyses the use of Poisson wavelet in more details.

Preisig came back to wavelets as modulating functions in Preisig [2010], where he discussed the suitability of input signal for further purpose of system identification. There is however no general interconnection of wavelets and modulating functions, but the utilizing multi-wavelets as modulating functions only.

3.1.4 Linear time-varying system identification

The identification of Linear Time-Varying (LTV) systems is the last application of wavelets in this chapter. It can arouse a discussion whether the LTV systems belong to the either linear or non-linear systems. In this thesis, we will consider LTV system as a specific extension of linear system.

When going through the number of publications concerning this theme, for example Chang and Shi [2010]; Dorfan et al. [2004]; Doroslovacki et al. [1998]; Ghanem and Romeo [1999]; Kougioumtzoglou and Spanos [2013]; Li et al. [2011]; Shan and Burl [2011]; Tsatsanis and Giannakis [1993]; Wei and Billings [2003]; Wei et al. [2008]; Xu et al. [2012], we can get the impression of existence the only one way of exploiting wavelets for LTV system identification. Namely, varying parameters are generally considered as n-dimensional functions so they can be expressed as a linear combination of suitable n-dimensional basis functions - wavelet functions in this case. Wavelets are used because of their superior selectivity in frequency domain, because of their inherent orthogonality and because of very large foundation of wavelet functions. Such an approach transforms time-varying system parameters into time-invariant coefficients of parameters approximation via wavelets. Therefore the whole problem becomes time-

invariant as well and is solvable as a classical linear system identification problem. This idea could be improved by using a shift-scale plane analysis Staszewski [1998], wherein the problem of the proper wavelet selection is addressed.

At first sight, the method could seem as quite simple compared to the obtained result. However, there is one big obstacle which the user should take care when employing this technique. The wavelet functions are (based on theory) square-integrable functions, therefore in discrete-time space even compactly supported functions. It consequently means that any their linear combination has also compact support, so the linear approximation of varying parameter as well. This is probably the biggest defect of the method and its main limiting factor for being used for prediction of evolution of system parameters. On the other hand, parameters approximation via wavelets provides us with an insight into the time-frequency parameters behavior and further analysis can disclose partial relations among parameters development and adjoining events. Moreover it is important to distinguish two points of view on parameters development: i) Parameter changes in order to ageing of the system. Such changes are irreversible, smooth and usually slow. ii) Quite opposite case is that parameter changes from its principle (e.g. a changeable weight inside a lift). These changes can be very quick, non-smooth and are definitely reversible.

The discussed method is demonstrated mainly on simulation case studies across the publications, but there are few of them applying it in practise, for instance in biomedical engineering for EEG analysis Li et al. [2011]; Wei et al. [2008] or on analysis of hysteretic behavior Chang and Shi [2010].

3.2 Wavelet transform in nonlinear system identification

The procedures of using wavelets on linear processes treated in the previous section are applicable for the identification of non-linear processes as well. There are few different methods of utilizing wavelets which, though regarded as general methods, belong rather into the non-linear section Sjoberg et al. [1995].

One of them is applying wavelet functions for a support vector machine algorithm as admissible support vectors. Wen et al. [2005a,b] propose a wavelet support vector machine with reproducing wavelet kernel especially for the identification of non-linear dynamics or approximating a non-linear function. The main advantages of using wavelets here are their compactness, orthogonality and a good reproducibility of wavelet kernel. Moreover, wavelet kernel usually performs much faster learning in comparison to standard neural networks or fuzzy logic Li and Liu [2006].

Another approach employs wavelet function as a sigmoid function within a neural network, then called wavelet network Zhang and Benveniste [1992]. A lot of publications have been published on this theme, for more details, the reader is referred to Adeli and Jiang [2006]; Billings and Wei [2005]; Ghanem and Romeo [2001]; Shi et al. [2005]; Wen et al. [2005a,b]; Zhang [1997].

3.3 Extensions of wavelet transform

3.3.1 Advanced wavelets

As time went on, the wavelet theory recorded further development and extension. For instance, new "families" of wavelet functions have been discovered and the wavelet theory has been established for non-orthogonal wavelet functions. Therefore, apart from the simple wavelet analysis, the bi-orthogonal wavelets Ho and Blunt [2003], wavelet frames (Sureshbabu and Farrell [1999]) or multi-wavelets Preisig [2010]; Strang and Strela [1994]; Strela and Strang [1994]; Strela and Vogan [1996] can be possibly used for system identification.

As the reader probably noticed, many wavelet applications have exploited specific class of wavelets, namely wavelets with an explicit mathematical description. Using of just these wavelets is not surprising. Since it is a common practise to use impulse or step functions (or their combination, e.g. pseudo-random binary signal) as a system input, the wavelet analysis of such signals can be mathematically derived only in case of wavelets with any mathematical expression. There is a different point of view as well - when the user has a possibility to design the system input, exploiting these wavelets can be advantageous due to the results in the form of direct mathematical formulae. The benefit is apparent: case study independent direct equations for parameters estimation. On the other hand, there are several wavelet families with no explicit expression, but much more advantageous from another point of view (see Chapter 4.5). Therefore it is not always desired to derive any general results, but rather to sketch a basic invention and derive its elementary (and general) properties like asymptotic behavior, reproducibility, etc.

The citation list is by no mean exhaustive since there are lot of papers published within many journals and it is almost impossible to track them all. Therefore the reader should take into consideration that the cited articles are only representatives of presented research directions that can be used further as starting points for searching details.

4 Wavelet transform

Basic concepts and ideas of WT were discussed within Chapter 1. This chapter introduces primers of WT, which is necessary for further work. It is widely known fact that there are several forms of FT according to the space where the transform is defined on. Analogously, since WT is an "extension" of FT, there are also several forms of WT, again - each defined on particular space of sequences or functions.

The Discrete Wavelet Transform (DWT) is a fundamental form of WT and is defined on $\ell^2(\mathbb{Z}_N)$, a Hilbert space of complex sequences of N-dimensional vectors, where the theory of the Discrete Fourier Transform (DFT) is built on as well. Although the wavelet theory is in general very complex, the simplicity of this form lies within the finite dimension of the space $\ell^2(\mathbb{Z}_N)$, which directly yields the completeness of the orthonormal wavelet basis.

Next form is defined on $\ell^2(\mathbb{Z})$, a Hilbert space of infinite, generally not periodic, complex signals. As it operates on discrete-time signals, the background theory is mostly adopted from the DWT. Main differences arise from the infinite dimensionality, what makes the theory more demanding, e.g. it is necessary to prove the completeness of the orthonormal (basis) set.

The third and the most complicated (from the theory point of view) form of WT is the CWT, which is defined on $L^2(\mathbb{R})$, a Hilbert space of complex square-integrable functions. Here, the construction of wavelets is usually reduced to the construction of a MultiResolution Analysis (MRA).

For purposes of further applying for real processes, only the DWT is worth to introduce in more details. The basic principle of other WT remains the same across different types of WT, however their background theories differ a lot and the reader can found more details in Frazier [1999]. The theories will not be presented within this thesis, but they will be used when necessary and an appropriate literature will be referred to. An overview of its well-known and established theory follows.

4.1 Prerequisities

Before further explanation of the DWT, let us recall some important preliminaries. See Frazier [1999]; Kolzow [1994] for further details on any of them.

- **Hilbert space and the best approximation of its element.** The DWT is defined on Hilbert spaces. Let $\mathcal{H}$ denotes Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Each Hilbert space has an orthonormal basis. If this orthonormal basis is countable, the Hilbert space is separable, see Rudin [1986]. Such a separable Hilbert spaces are for example aforementioned $\ell^2(\mathbb{Z}_N)$, $\ell^2(\mathbb{Z})$ and $\ell^2(\mathbb{R})$ spaces.

  Let $\mathcal{M}$ be some closed subspace in $\mathcal{H}$: $\mathcal{M} \subset \mathcal{H}$ and let $\mathcal{M}^\perp$ be its orthogonal complement: $\mathcal{M}^\perp = \{x \in \mathcal{H} |\ \langle x, y \rangle = 0\ \forall y \in \mathcal{M}\}$. We can write $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$ and the best approximation x_{app} of any vector $x \in \mathcal{H}$ is then its orthogonal projection onto subspace $\mathcal{M}$. More precisely, $x_{app} \in \mathcal{M}$ is uniquely determined by the condition $x - x_{app} \in \mathcal{M}^\perp,\ x \in \mathcal{H}$.

- **Periodic extension of a vector.** For the sake of simple notation, it is worth to make a convention by defining a periodic extension of a vector to be defined at all integers (similar as within the DFT theory). Let us define the periodic extension of vector $z = [z(0), z(1), \ldots, z(N-1)] \in \ell^2(\mathbb{Z}_N)$ as $z(t+N) = z(t)\ \forall t \in \mathbb{Z}$.

- **Discrete Fourier transform.** A standard notation of the DFT is used across the thesis, i.e.

$$\hat{z}(\omega) = \mathcal{F}\{z(t)\} = \sum_{t=0}^{N-1} z(t) \cdot e^{-2\pi i \omega t/N}. \tag{4.1}$$

  For further purposes of this work recall two important equations from DFT theory which holds for any $z, w \in \ell^2(\mathbb{Z}_N)$:

  - Parseval's theorem:

$$\langle z, w \rangle = \frac{1}{N} \langle \hat{z}, \hat{w} \rangle \tag{4.2}$$

  - Plancherel's formula

$$||z||^2 = \frac{1}{N} ||\hat{z}||^2 \tag{4.3}$$

- Furthermore, some additional operators are defined Frazier [1999].

  1. **Definition 4.1: Translation operator:** Suppose $z \in \ell^2(\mathbb{Z}_N)$ and $k \in \mathbb{Z}$. Then define

$$R_k z(t) = z(t-k) \text{ for } t \in \mathbb{Z}. \tag{4.4}$$

     R_k is called the *translation by* k operator and $R_k z$ is called *translate of* z *by* k.

2. **Definition 4.2: Convolution of two vectors:** For $z, w \in \ell^2(\mathbb{Z}_N)$, the convolution $z * w \in \ell^2(\mathbb{Z}_N)$ is the vector with components

$$z * w(\omega) = \sum_{t=0}^{N-1} z(\omega - t)w(t), \quad \forall \omega. \tag{4.5}$$

3. **Definition 4.3: Conjugate reflection:** For any $z \in \ell^2(\mathbb{Z}_N)$, define $\tilde{z} \in \ell^2(\mathbb{Z}_N)$ by

$$\tilde{z}(t) = \overline{z(-t)} = \overline{z(N-t)}, \quad \forall t. \tag{4.6}$$

We call $\tilde{z}$ the *conjugate reflection of* z.

Moreover, the following properties indispensable for discrete wavelet theory hold:

$$z * w(k) = \langle z, R_k \tilde{w} \rangle, \tag{4.7}$$
$$z * \tilde{w}(k) = \langle z, R_k w \rangle. \tag{4.8}$$

4.2 Multiresolution analysis

The main idea of the DWT is to decompose original Hilbert space $\ell^2(\mathbb{Z}_N)$ into two subspaces, space of approximations $\mathcal{V}$ and space of details $\mathcal{W}$. This is known as analysis at the 1^{st} level. Since the (Hilbert) space of approximations is closed and has countable basis, it is separable and can be decomposed further. In general, this is known as analysis at the p^{th} level. Let $\mathcal{V}_j$ and $\mathcal{W}_j$ denote space of approximations at the j^{th} level and space of details at the j^{th} level, respectively. Then we can write

$$\begin{aligned} \ell^2(\mathbb{Z}_N) &= \mathcal{V}_1 \oplus \mathcal{W}_1, \\ \mathcal{V}_1 &= \mathcal{V}_2 \oplus \mathcal{W}_2, \\ &\vdots \\ \mathcal{V}_{p-1} &= \mathcal{V}_p \oplus \mathcal{W}_p, \end{aligned}$$

or in more compact form

$$\ell^2(\mathbb{Z}_N) = \mathcal{V}_p \oplus \mathcal{W}_p \oplus \mathcal{W}_{p-1} \oplus \cdots \oplus \mathcal{W}_1. \tag{4.9}$$

Both subspaces $\mathcal{V}_j, \mathcal{W}_j$ for some j have the same dimension, what consequently means that spaces $\mathcal{V}_j, \mathcal{W}_j$ have dimension $N/2^j$. That is the reason, why N has to be dividable by 2^p, where p is maximum level of analysis. Note that the nomenclature *space of approximations* and *space of details* has its origin in two essential wavelet properties. The first of them is the possibility to construct the best approximation of any vector by orthogonal projection (what the DWT does) and the second is, concerning the wavelets on a particular level of analysis, mutual exclusivity of wavelets in a frequency domain. This will be discussed more in details in Section 4.4.

4.3 Discrete wavelet transform

This section gives a brief sketch of the DWT theory. Full DWT theory can be found in Frazier [1999]; Kolzow [1994] together with proofs of propositions stated here.

Let $N = 2M$ and $\varphi, \psi \in \ell^2(\mathbb{Z}_N)$. Then the set $B = \{R_{2k}\varphi\}_{k=0}^{M-1} \cup \{R_{2k}\psi\}_{k=0}^{M-1}$ is orthonormal basis in $\ell^2(\mathbb{Z}_N)$ if and only if the matrix

$$A(n) = \frac{1}{\sqrt{2}} \begin{bmatrix} \hat{\varphi}(t) & \hat{\psi}(t) \\ \hat{\varphi}(t+M) & \hat{\psi}(t+M) \end{bmatrix} \tag{4.10}$$

is unitary for all $t = 0, \ldots, M-1$. Such an orthonormal basis B is called wavelet basis at the 1^{st} level, and vectors φ, ψ are called its generators. Vector φ is called a father wavelet and vector ψ is called a mother wavelet. Furthermore, let φ be a vector such that a set $\{R_{2k}\varphi\}_{k=0}^{M-1}$ is orthonormal. Thereafter one of the possible vector ψ could be constructed as follows:

$$\psi(n) = (-1)^t \overline{\varphi(1-t)}, t = 1, \ldots, N.$$

Let us have a signal $z \in \ell^2(\mathbb{Z}_N)$. Its coefficients in the basis B can be expressed as the inner products of z with the basis vectors. By taking advantage of the equation (4.8), vector z in the basis B can be written as

$$[z]_B = \left[z * \tilde{\varphi}(0), z * \tilde{\varphi}(2), \ldots, z * \tilde{\varphi}(N-2), z * \tilde{\psi}(0), z * \tilde{\psi}(2), \ldots, z * \tilde{\psi}(N-2) \right].$$

Definition 4.4: Let $z \in \ell^2(\mathbb{Z}_N)$, $w \in \ell^2(\mathbb{Z}_M)$ and $N = 2M$. Define a downsampling operator $D : \ell^2(\mathbb{Z}_N) \to \ell^2(\mathbb{Z}_M)$, $D(z(t)) = z(2t)$ for $t = 0, \ldots, M-1$ and an upsampling operator $U : \ell^2(\mathbb{Z}_M) \to \ell^2(\mathbb{Z}_N)$, $U(w(t)) = w\left(\frac{t}{2}\right)$ for even t and $U(w(t)) = 0$ for odd t.

These operators possess following properties:

$$D(U(z)) = z, \tag{4.11}$$

$$U(D(z)) = \frac{1}{2}(z + z^*), \tag{4.12}$$

where $z^*(t) = (-1)^t z(t)$. $[z]_B$ could be now expressed as

$$[z]_B = \left[D(z * \tilde{\varphi}), D(z * \tilde{\psi}) \right]. \tag{4.13}$$

(4.13) illustrates a representation of the signal z by the vectors of approximations and details and it is also in accordance with the idea of the best approximation. To construct the wavelet bases for the whole analysis at the p^{th} level seems to be very complex. However, the wavelet basis has one important property, that it can be constructed only from generators at the 1^{st} level. This will be described in the following proposition.

Proposition 4.1: *Let N be divisible by 2^p, $p \in \mathbb{N}$ and let $\varphi_1, \psi_1 \in \ell^2(\mathbb{Z}_N)$ is a pair of generators of wavelet basis at the 1^{st} level. Let us construct the sequence of pairs of vectors φ_j, ψ_j, $j = 2, \ldots, p$ as follows:*

$$\varphi_j(t) = \sum_{k=0}^{2^{j-1}-1} \varphi_j\left(t + \frac{kN}{2^{j-1}}\right), \ \psi_j(t) = \sum_{k=0}^{2^{j-1}-1} \psi_j\left(t + \frac{kN}{2^{j-1}}\right), \ t = 0, \ldots, \frac{N}{2^{j-1}} - 1.$$

Then the set $\{\varphi_j, \psi_j\}_{j=1}^p$ is a sequence of wavelet basis generators for analysis up to the p^{th} level.

The wavelet analysis is now performed in accordance to the MRA. Firstly, the analysis (4.13) of $z \in \ell^2(\mathbb{Z}_N)$ is done to obtain vectors of approximation $a_1 = D(z * \tilde{\varphi_1})$ and details $d_1 = D(z * \tilde{\psi_1})$ on the 1^{st} level, both from $\ell^2(\mathbb{Z}_{N/2})$. Then continue with analysis of the vector a_1 by repeating the procedure with φ_2 and ψ_2, so we obtain vectors $a_2 = D(a_1 * \tilde{\varphi_2})$ and $d_2 = D(a_1 * \tilde{\psi_2})$, both from $\ell^2(\mathbb{Z}_{N/2^2})$. The procedure continues up to the p^{th} level, where we have final approximation a_p and final details d_p, both from $\ell^2(\mathbb{Z}_{N/2^p})$.

Definition 4.5: Let N is divisible by 2^p, $p \in \mathbb{N}$ and let $\varphi_j, \psi_j \in \ell^2(\mathbb{Z}_{N/2^{j-1}})$ is a pair of generators of a wavelet basis of $\ell^2(\mathbb{Z}_{N/2^{j-1}})$, $j = 1, \ldots, p$. Then the vector $[d_1, d_2, \ldots, d_p, a_p]$, obtained as described above, is called an analysis of the vector $z \in \ell^2(\mathbb{Z}_N)$ at the p^{th} level.

So far we treated only the wavelet decomposition. Let us now have a brief look at the reconstruction.

Proposition 4.2: *Let $M \in \mathbb{N}, N = 2M$ and let $\varphi_1, \psi_1 \in \ell^2(\mathbb{Z}_N)$ be the generators of wavelet basis of $\ell^2(\mathbb{Z}_N)$. Let $D(z * \tilde{\varphi_1}) = x_1 \in \ell^2(\mathbb{Z}_M)$, $D\left(z * \tilde{\psi_1}\right) = y_1 \in \ell^2(\mathbb{Z}_M)$. Then the following equation holds*

$$\varphi_1 * U(x_1) + \psi_1 * U(y_1) = z, \ \forall z \in \ell^2(\mathbb{Z}_N). \tag{4.14}$$

This is an equation of the perfect reconstruction of the original signal $z \in \ell^2(\mathbb{Z}_N)$. For the reconstruction we can then write analogously

$$\begin{aligned}
\varphi_p * U(a_p) + \psi_p * U(d_p) &= a_{p-1}, \\
&\vdots \\
\varphi_j * U(a_j) + \psi_j * U(d_j) &= a_{j-1}, \\
&\vdots \\
\varphi_1 * U(a_1) + \psi_1 * U(d_1) &= z.
\end{aligned}$$

4.4 Frequency properties

Recall that in order to the theory built on the $\ell^2(\mathbb{Z}_N)$ space, discrete frequency domain is proposed. Due to the practical applications we consider a real space signal z sampled by $T_s = \frac{1}{f_s}$ with sampling frequency f_s. Such a signal has full and symmetric (double sided) spectra. To comply with Shannon-Kotelnik theorem we count only single sided spectra, however, due to the necessity of retaining of the full signal energy, we have to multiply it by 2 in accordance to discrete wavelet theory. If we express DFT of wavelet analysis in (4.13), we obtain

$$\mathcal{F}\{[z]_B\} = \left[\frac{1}{4}\hat{z}\left(\frac{\omega}{2}\right)\hat{\tilde{\varphi}}\left(\frac{\omega}{2}\right) \ , \ \frac{1}{4}\hat{z}\left(\frac{\omega}{2}\right)\hat{\tilde{\psi}}\left(\frac{\omega}{2}\right)\right]. \tag{4.15}$$

what can be repeated up to the p^{th} level. That is why the wavelet analysis can be thought of as a filtering operation and it is the reason why φ_j, ψ_j are sometimes called wavelet filters. Moreover, this is why the subspaces are called after approximations and details. It is logical that the approximation $D(z * \tilde{\varphi})$ of the original signal z contains the smallest frequencies and the details $D(z * \tilde{\psi})$ contain the highest frequencies, so $\hat{\tilde{\varphi}}\left(\frac{m}{2}\right)$ is low-pass filter and $\hat{\tilde{\psi}}\left(\frac{\omega}{2}\right)$ is high-pass filter.

The frequency characteristics of both types of wavelet filters cover full range of frequencies, however, in practice, only that half of the frequency range can be considered which has a major influence. We say that a wavelet function is localised in some frequency interval if most of its components outside that interval are 0 or relatively small. Moreover, the frequency localisation of mother and father wavelets are almost exclusive. While the father wavelet is localised over the first half of the frequency domain (the half with low frequencies), the mother wavelet covers the half with high frequencies, see upper characteristics for level 1 in Figure 4.1. Based on a metaphor, father wavelet is also sometimes called "filter" and mother wavelet is called just "wavelet". Numerically, at the 1^{st} level analysis, the full frequency range $1 \cdot \frac{f_s}{N}, 2 \cdot \frac{f_s}{N}, \ldots, \frac{N}{2} \cdot \frac{f_s}{N}$ is divided into two bands: $\frac{f_s}{N}, \ldots, \frac{N}{4} \cdot \frac{f_s}{N}$ for the low-pass filter (father wavelet) and $\frac{N}{4} \cdot \frac{f_s}{N}, \ldots, \frac{N}{2} \cdot \frac{f_s}{N}$ for the high-pass filter (mother wavelet).

A consequence of joining the MRA and the equation (4.15) is that the procedure can be applied repeatedly up to the p^{th} level. Since the transform is performed according to the (4.13), the space of approximations at the 1^{st} analysis level has dimension $N/2$ (due to the downsampling operator), what consequently means that the highest distinguishable frequency is also half. Therefore, at the next analysis level, the wavelets frequency localisation look like that in Figure 4.1 for level 2. The Figure 4.1 demonstrates the repeating of the filtering procedure within the DWT for higher levels. The overview of resulting wavelet filters and their frequency localisations is in Table 4.1 and their examples are depicted in Figure 4.2.

As can be seen from the figure Figure 4.2, wavelet filters divide the frequency domain equidistantly in logarithmic coordinates. There is one more option of frequency domain division introduced by a discrete wavelet packet transform (DWPT). In DWPT, the decomposition is performed not only on the approximations, but on the details as well.

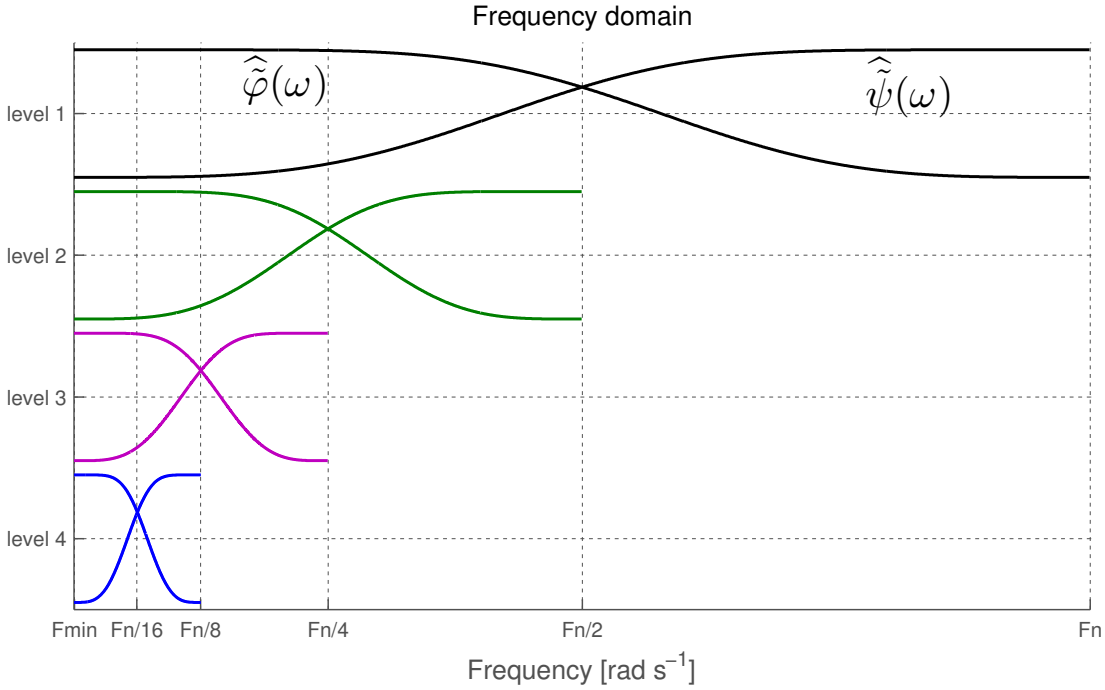


Figure 4.1: Frequency domains of wavelet filters at distinct analysis levels.

Therefore, final division of full frequency range is equispaced in linear coordinates Gao and Yan [2011]; Mohallem and Kawakami [2006]. Note that whether some particular frequency (important from the application point of view) belongs (or not) into some specific frequency band given by a wavelet function depends on the sampling time as well as on the total length of the analysed signal.

Table 4.1: Table of wavelet filters and their main intervals at distinct levels.

Analysis	**localisation**		**Frequency-domain filter**
	f_{min}	f_{max}	
1^{st} level details	$\frac{f_s}{4}$	$\frac{f_s}{2}$	$\hat{\tilde{\psi}}\left(\frac{\omega}{2}\right)$
2^{nd} level details	$\frac{f_s}{8}$	$\frac{f_s}{4}$	$\hat{\tilde{\psi}}\left(\frac{\omega}{4}\right)\cdot\hat{\tilde{\varphi}}\left(\frac{\omega}{2}\right)$
3^{rd} level details	$\frac{f_s}{16}$	$\frac{f_s}{8}$	$\hat{\tilde{\psi}}\left(\frac{\omega}{8}\right)\cdot\hat{\tilde{\varphi}}\left(\frac{\omega}{4}\right)\cdot\hat{\tilde{\varphi}}\left(\frac{\omega}{2}\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
p^{th} level details	$\frac{f_s}{2^{p+1}}$	$\frac{f_s}{2^p}$	$\hat{\tilde{\psi}}\left(\frac{\omega}{2^p}\right)\cdot\prod_{j=1}^{p-1}\hat{\tilde{\varphi}}\left(\frac{\omega}{2^j}\right)$
p^{th} level approximation	0	$\frac{f_s}{2^{p+1}}$	$\prod_{j=1}^{p}\hat{\tilde{\varphi}}\left(\frac{\omega}{2^j}\right)$

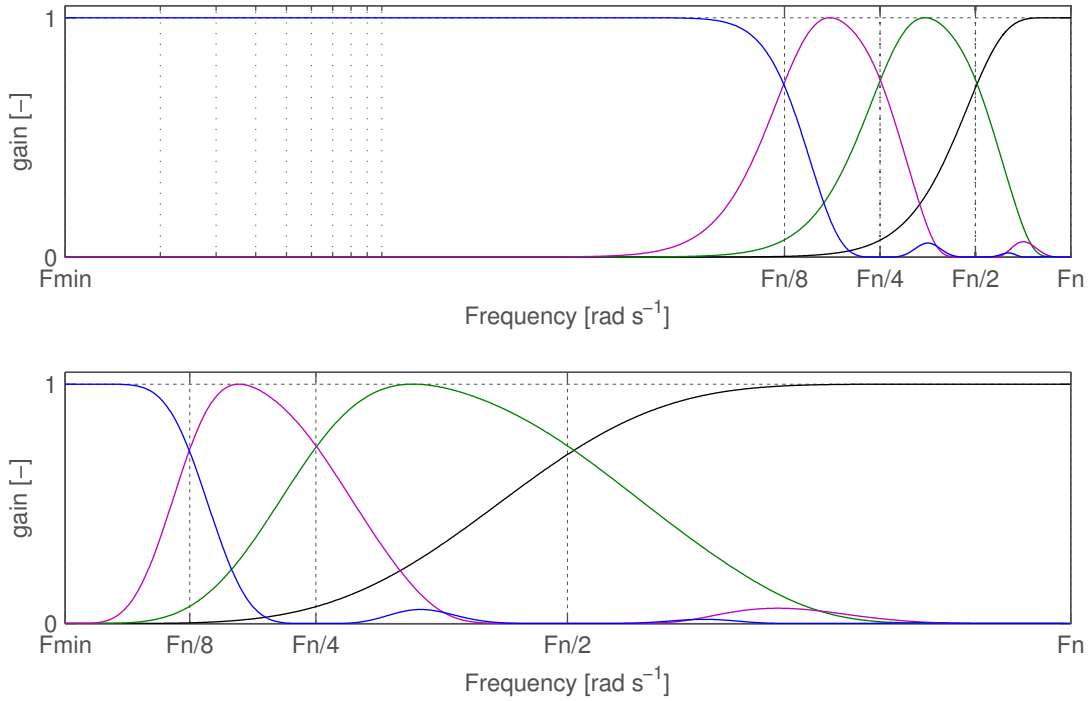


Figure 4.2: Resulting frequency domains of wavelet filters.
Both figures depict the same wavelet filters, the upper in logarithmic frequency and the lower in linear frequency.

The sense of the wavelet function in both time and frequency domains is depicted in Figure 4.3. Let us have the wavelet db8 at the 5^{th} level from the Daubechies wavelet family (see next Section 4.5.1) and analyse the data sampled by $f_s = 3$ Hz. The upper figure depicts the spectra of the wavelet filter (blue line). The horizontal violet line shows the amplitude of -10 dB and vertical violet lines stand for limit frequencies of the wavelet filters for the gain of -10 dB. Note that this gain has been chosen for the purpose of the example, however, as the figure shows, the wavelet filter's localisation in frequency can be considered much wider. The important fact is that the wavelet filter accentuates those frequencies between violet ones while the others are suppressed significantly. In this example, the violet frequencies are 0.09 Hz and 0.21 Hz. Moreover, two frequencies have been chosen to show the analogy in time: 0.14 Hz (red color) and 0.3 Hz (green color). Two lower plots depict the wavelet function together with the harmonic function of appropriate frequency. The similarity in time can be observed for the frequency 0.14 Hz (Figure 4.3 down left) what results in large coefficient of wavelet transform. Since the wavelet is shifted during transform, a significance of the wavelet coefficient is marked by its absolute value which obviously does not depend on the shift. Right opposite case is the frequency 0.3 Hz (Figure 4.3 down right) which is "too fast", thus is finally suppressed. The wavelet coefficients for this case are 0.0334 for 0.14 Hz and $1.13 \cdot 10^{-7}$ for 0.3 Hz.

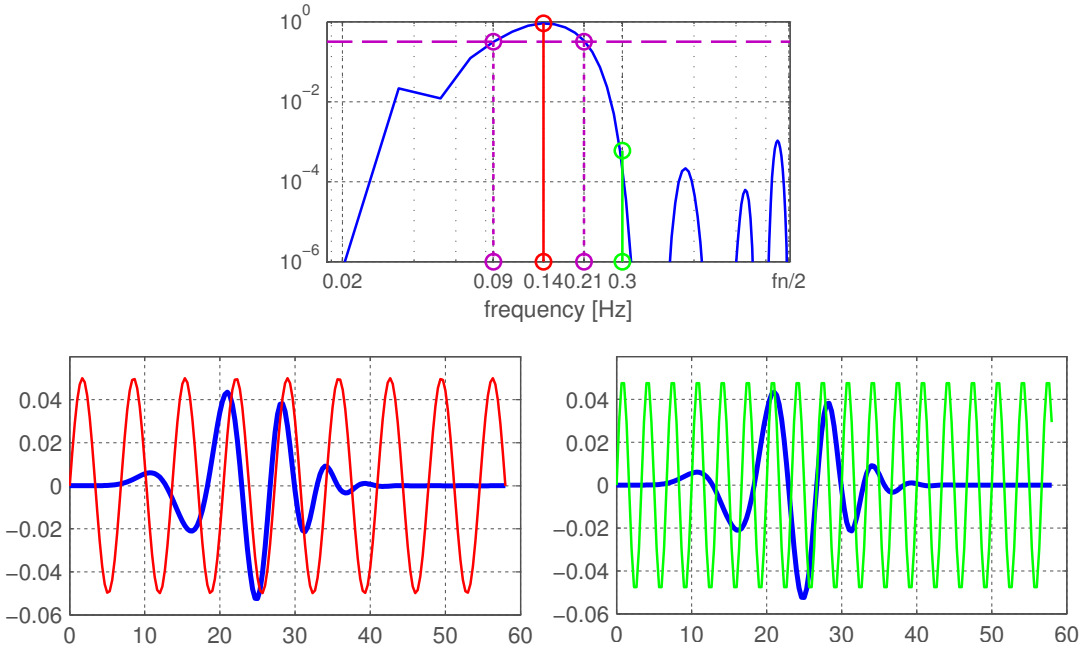


Figure 4.3: Time-frequency analogy of wavelet transform.
The meaning of colors is preserved in all figures.

4.5 Basic wavelets and their properties

Several different wavelet functions have been developed. More precisely, several different methods to create or compute wavelet functions have been developed, each of them producing a set of wavelet functions. Each of these methods has been proposed in order to find a set of wavelet functions complying some particular property usually specified by a parameter, e.g. length of wavelet filter support. These sets are called *wavelet families*. This section starts with a list of basic wavelets properties and then introduces particular wavelet families classifying them from those properties point of view, not considering the type of wavelet analysis.

A suitability of using of particular wavelet family varies in accordance to properties of its wavelets and also in accordance to the application itself. The main qualitative parameters of wavelets are:

1. **Wavelet filter support:** Compact (finite) support is with no doubts one of the most desired wavelets properties because of two reasons: Firstly, a compact support results in a finite multiplication in the DWT, what consequently yields to simple practical implementation of wavelet analysis. Since there is no need to approximate the wavelet function (to have a compact support), all computations are exact. The second reason is a good time localisation of the wavelet, what will be discussed in more details later.

2. **Symmetry and antisymmetry:** Since both the scaling function and the mother wavelet can be seen as filters, their phase characteristic plays an important role as the negative derivative of the phase is a group delay. It is very desirable in signal processing application that the filters have a linear phase, thus constant group delay. Even in image processing this issue is known as very crucial in the restoration process Ahuja et al. [2005]. The linear phase of wavelet filter requires its scaling function to be symmetric or antisymmetric.

3. **Orthogonality:** This property guarantees the independence of wavelets in time (their shifts and scales), thus some specific kind of exclusivity in signal analysis. Note that there must always be a trade-off between the orthogonality and symmetry of wavelets, since it is not possible to impose both of them simultaneously. One of the big issues in research was to design orthogonal wavelets which are symmetric to the degree possible. Some of results are e.g. symlets Daubechies et al. [1992]. It is also possible to exploit wavelets which are non-orthogonal, nevertheless, the wavelet theory is then much complicated and is built up on theory of Riesz's bases. Moreover, the wavelet synthesis is not perfect anymore, since there is always an error between original and reconstructed signal.

4. **Number of vanishing moments:** Some wavelets suppress moment functions, thus polynomial functions of certain order as well. It results in more sparse representation of the wavelet analysis, what can significantly save the memory space while implemented Fugal [2009]. Again, there is a trade-off between number of vanishing moments and the length (dimension) of the wavelets.

5. **Existence of scaling function:** The simple rule holds about this property: When the scaling function does not exist, the analysis is not orthogonal. Note that the rule is rather theoretical and has no impact on the choice of wavelet family to be applied in practise, since the (non)orthogonality property of particular wavelet family is always known.

6. **Expression:** An explicit representation of the wavelet will generally result in faster computation of their elements as well as coefficients of wavelet analysis. However, the most of explicitly formulated wavelets have infinite support except two of them: Haar wavelet and B-spline wavelets. Note that the infinity support is usually due to the exponential function in the expression, hence these wavelets need not be well localised in time.

7. **Time-frequency localisation:** A good localisation of wavelet in either time or frequency determines the possibility of the wavelet to detect particular phenomena in corresponding domain. However, in order to the uncertainty principle, both the compact support and the band-limitations property cannot be attained simultaneously. In other words, the more strict resolution in frequency, the larger support of wavelet, thus the worse time resolution. Analysis in time proceeds by shifting the wavelet along the time axis and analysis in frequency proceeds by

scaling the wavelet. Obviously, widening the wavelet shifts its frequency content towards low frequencies and conversely, narrowing the wavelet shifts its frequency content towards high frequencies.

4.5.1 Wavelet families

The wavelets are to be chosen to fit the application Ahuja et al. [2005]. It would not be reasonable to pick one particular wavelet, use it for all applications and always expect a good results. Different applications require different properties of wavelets. In the following, an overview of wavelet families is provided with emphasis on their main properties. For more details see Fugal [2009].

4.5.1.1 Orthogonal wavelets

The property of orthogonality has already been discussed in Section 4.5, however there is one fact which has not been mentioned - the orthogonality is an essential presumption in wavelet theory in general. Orthogonal wavelets enable orthogonal MRA as well as the possibility of perfect reconstruction. When orthogonal wavelets are used, the energy of the signal is preserved in both time and frequency domains by virtue of Parseval's theorem (4.2). Moreover, most of orthogonal wavelets have a compact support, what makes them the most often used wavelets in the DWT, hence in practise.

1. **Daubechies wavelets** are called in honor of its inventor, well known mathematician Ingrid Daubechies. They are usually denoted as db$N$, where N is the order. The order N actually determines the wavelet properties: the wavelet db$N$ has support of the length of $2N$ and has N vanishing moments[1]. The wavelets are not symmetric and do not have an explicit expression. The only exception is a Haar wavelet, what is a special case of Daubechies wavelet for $N = 1$. The Haar wavelet resembles the step function and is the most simple wavelet, what makes it the most used wavelet. Note that since the short support, it has an excellent time localisation, but at the expense of the resolution in frequency.

2. **Symlet wavelets** have also been invented by Ingrid Daubechies. Symlets are more symmetric than Daubechies wavelets, what is their main advantage. The wavelets are denoted as sym$N$ and have the same property as db$N$, what is the length of $2N$ and N vanishing moments.

3. **Coiflet wavelets**, despite their name, are also made by Ingrid Daubechies at the request of Ronald Coifman. They are denoted as coif$N$ and have $2N$ (mother wavelet) or $2N - 1$ (father wavelet) vanishing moments and length of $6N - 1$. So when looking at the support, coiflets have less number of vanishing moments than Daubechies wavelets or symlets.

[1] In literature, dual notation can be found. One where db$N$ has the length of $2N$ and one where db$N$ has the length of N.

4.5.1.2 Crude wavelets

On a contrary to orthogonal wavelets, crude wavelets have usually explicit expression, therefore are very smooth, but also have an infinite support. Such an expression enables a very good localisation in frequency. Crude wavelets can only be used in the CWT. Although there is always a possibility to evaluate their explicit formula in equispaced points in time to obtain a discrete-time approximation (with so called "effective support"), such an approximation is not orthogonal and can not be used in the DWT.

1. **Shannon wavelets** is the dual wavelet to Haar, since it is defined as a rectangular bandpass in frequency domain, thus have a perfect resolution and localisation in frequency. On the other hand, it leads to the *sinc* function in time domain, hence very poor time localisation. It has infinite number of vanishing moments, is infinitely differentiable and its integer shifts are orthogonal to each other.

2. **Morlet wavelets** has already been introduced in Chapter 1 as a (complex) harmonic function multiplied by Gaussian window. The wavelet is symmetrical and has an effective support from -4 to $+4$.

3. **Gaussian wavelets** are derived as derivatives of a single Gaussian function e^{-t^2}. They are either symmetric or antisymmetric, have a very simple expression and have an infinite number of vanishing moments. Low-order (small number of derivatives) Gaussian wavelets have quite good time resolution.

4. **Mexican hat** is the most known representative of Gaussian wavelet. It is computed as its negative second derivative. Its effective support is from -5 to $+5$ and the wavelet is symmetric. With very narrow localisation and very rapid decay in time and since the human eye work somewhat like a Mexican hat Fugal [2009], it is an unique choice for vision analysis.

5. **Meyer wavelets** are, as well as Shannon wavelet, defined in frequency domain, but in contrast to Shannon wavelet, sharp edges in frequency are "replaced" by smooth function, what causes faster decay in time domain. Its effective support is from -8 to $+8$.

6. **Bi-orthogonal wavelets** were obtained as the result of both symmetry and perfect reconstruction requirements. As those properties are incompatible in case of just one wavelet filter, bi-orthogonal wavelets have different filters for decomposition and reconstruction.

5 Introduction to linear time-invariant system identification

This chapter briefly provides the reader with basics from the area of linear time-invariant systems, design of the model structure and its parameters identification via prediction error method. The main focus is, however, laid on the methods and principles, which are to be used further in the thesis.

5.1 Linear time-invariant systems

LTI systems is the most important class of dynamic systems exploited in both theory and practice. Since the most of real processes record some non-linearity, LTI systems usually serve as an approximation of the processes encountered in real life. Moreover, their simple structure is often justified and results based on them prove sufficient for the application. As the theory of linear systems is well-established and widely known, only the main terminology is collected in this section. Moreover, for the purposes of this thesis, the main focus will be laid on discrete-time systems. For further details see e.g. Ljung [1999]; Van den Hof [1996].

A LTI causal continuous-time system can be described through its time-invariant impulse response $g(t)$ as

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau,\ t \in \mathbb{R},$$

where $u(t), y(t)$ are system input and output signals, respectively. Its discrete-time equivalent is

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k),\ t \in \mathbb{N}$$

with $g(k)$ being uniformly sampled impulse response. Analogously, the effect of disturbances can be expressed as

$$v(t) = \sum_{k=0}^{\infty} h(k)e(t-k)$$

with $e(t)$ being a zero-mean white noise with variance σ_e^2 and general LTI system can then be written as

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t).$$

By defining the forward shift operator q as $qu(t) = u(t+1)$, we can rewrite a general LTI system as

$$y(t) = G(q)u(t) + H(q)e(t), \tag{5.1}$$

where

$$G(q) = \sum_{k=1}^{\infty} q^{-k}g(k) \ , \ H(q) = \sum_{k=0}^{\infty} q^{-k}h(k) \tag{5.2}$$

are transfer functions of appropriate linear systems, $G(q),\ H(q) \in \mathcal{K}_2$, where $\mathcal{K}_2$ stands for a set of complex functions squared integrable on the unit circle and analytic in the exterior of the unit circle. When $H(q)$ is a monic transfer function, then moreover $H^{-1}(q) \in \mathcal{K}_2$. Monicity implies that $h(0) = 1$ and we can therefore write

$$H(q) = 1 + \sum_{k=1}^{\infty} h(k)q^{-k}.$$

Throughout the rest of the thesis, a general discrete-time LTI system is considered to be described by (5.1). Note that complex number $G(\omega)$ is the value of the transfer function $G(q)$ evaluated in the point $q^{-1} = e^{i\omega}$ and it bears the information about the steady state behavior of the system when the input is harmonic function with frequency ω.

For predicting the behavior of the LTI system (5.1), denote the conditional expectation on the future output in time t based on knowledge of outputs up to the time $t-1$ as $\hat{y}(t|t-1)$. Note that when predicting the output in time t, the input $u(t)$ is already known, so the predictor describes how the system responses on the input in time t. Then, a general one step ahead predictor of the LTI system (5.1) is (Ljung [1999])

$$\hat{y}(t|t-1) = \left[1 - H^{-1}(q)\right] y(t) + H^{-1}(q)G(q)u(t). \tag{5.3}$$

5.2 Models of linear time-invariant systems

The first step in a way of a System IDentification (SID) is to determine the structure of the model to be identified. In other words, the procedure begins with choosing the set or class of models in which the most suitable model is searched. Such a choice of the model structure depends on several things concerning the knowledge of the system of interest like its inner physical structure, order of the system or its approximation or even the fact that the only knowledge of the system we have is the numbers of inputs and outputs. Whatever the system to be modelled is, a suitable classification of its model structure is very important from the identifiability point of view and directly affects the result of SID process.

Let us recall a basic notation used in SID theory. Let $\mathcal{M}$ denotes the set of all models with particular structure and let $\theta \in \mathbb{R}^d$ stands for the vector of unknown parameters of the model. Obviously, values of parameters are not arbitrary numbers, but have to satisfy some constraints given for example by physics. Hence we can formally write $\theta \in D_{\mathcal{M}} \subset \mathbb{R}^d$, where $D_{\mathcal{M}}$ is set of all admissible vectors of parameters. Then $\mathcal{M}(\theta),\ \theta \in D_{\mathcal{M}}$ is one specific admissible model from the whole set $\mathcal{M}$ (Ljung [2002]).

Now, each model is described by the structure and by parameters, which are usually unknown and are the main values to be estimated. A general transfer function model of LTI system can now be formulated as

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \tag{5.4}$$

with $G(q,\theta), H(q,\theta)$ being transfer functions with appropriate model structure $\mathcal{M}$. Realize that (5.4) describes a set of all possible models. In order to (5.3), the one step ahead predictor of such models is

$$\hat{y}(t|t-1,\theta) = \left[1 - H^{-1}(q,\theta)\right] y(t) + H^{-1}(q,\theta)G(q,\theta)u(t). \tag{5.5}$$

There is a number of distinct model structures which are frequently used and which are sufficient enough to cover most of applications. The main representatives are state space models for cases where an internal structure of the system is known and transfer function models taking only the input-output behavior of the system into consideration. The most known representatives of the latter model structures are Auto-Regressive with eXternal input (ARX), Auto-Regressive Moving Average with eXternal input (ARMAX), Output Error (OE) or the most general Box Jenkins (BJ) structures. For the purposes of this thesis, we focus on the ARX model. More details can be found in Ljung [1999].

The ARX structure represents a simple input-output relation and is described as a single linear difference equation

$$y(t) + \sum_{k=1}^{n_a} a_k y(t-k) = \sum_{k=0}^{n_b} b_k u(t-k) + e(t). \tag{5.6}$$

Analogously to the (5.2) and to the (5.4), we can write

$$A(q,\theta) = 1 + \sum_{k=1}^{n_a} q^{-k} a_k\ ,\ B(q,\theta) = \sum_{k=0}^{n_b} q^{-k} b_k, \tag{5.7}$$

where the parameters vector θ comprises of the unknown parameters $a_k,\ b_k$. The model structure is then

$$A(q,\theta)y(t) = B(q,\theta)u(t) + e(t). \tag{5.8}$$

The term $A(q,\theta)y(t)$ represents an auto-regressive part and the term $B(q,\theta)u(t)$ represents an exogenous input. When comparing with the linear model (5.4), we can see that

$$G(q,\theta) = \frac{B(q,\theta)}{A(q,\theta)}\ ,\ H(q,\theta) = \frac{1}{A(q,\theta)}. \tag{5.9}$$

For a special case of $n_a = 0$, the ARX model structure becomes a Finite Impulse Response (FIR) filter.

5.3 Prediction error method

Prediction Error Methods (PEM) is a broad family of parameter estimation methods, which have a close relationship with the maximum likelihood method (Ljung [2002]). PEM is based on (as the name marks) the minimization of the prediction error, namely one step ahead prediction error. The finally identified model is then the best one step ahead predictor from the set of all admissible models $\mathcal{M}(\theta),\ \theta \in D_{\mathcal{M}}$.

It is assumed that discrete input-output signals are generated by the system

$$y(t) = G_0(q)u(t) + H_0(q)e(t), \tag{5.10}$$

where $(\bullet)_0$ describes the true system transfer functions, and that the model is described by (5.4). The prediction of the model is (5.5) and prediction error is

$$\begin{aligned}\varepsilon(t,\theta) &= y(t) - \hat{y}(t|t-1,\theta)\\ &= y(t) - \left\{\left[1 - H^{-1}(q,\theta)\right] y(t) + H^{-1}(q,\theta)G(q,\theta)u(t)\right\}\\ &= H^{-1}(q,\theta)y(t) - H^{-1}(q,\theta)G(q,\theta)u(t)\\ &= H^{-1}(q,\theta)\left[y(t) - G(q,\theta)u(t)\right].\end{aligned} \tag{5.11}$$

The filtered prediction error is

$$\varepsilon_F(t,\theta) = F(q)\varepsilon(t,\theta) \tag{5.12}$$

with $F(q)$ being a stable linear filter.

In general, the problem of parameters estimation is an optimization task:

$$\arg\min_{\theta \in D_{\mathcal{M}}} \left[V(\mathcal{M}(\theta), \mathcal{D})\right]. \tag{5.13}$$

The optimization proceeds over the set of admissible parameters $\theta \in D_{\mathcal{M}}$ and the criterion function V to be minimized is some scalar-valued function of the data $\mathcal{D}$ and model structure $\mathcal{M}$, which evaluates the quality of the model. The criterion function is usually marked as $V_N(\theta)$, N denotes the number of data points from which the estimate is computed. For PEM, the criterion function is generally taken as

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \ell\left(\varepsilon_F(t,\theta)\right), \tag{5.14}$$

where $\ell(\bullet)$ is a positive scalar-valued function and is taken as $\ell(x) = \frac{1}{2}x^2$ very often. Finding of model parameters can then be formulated as

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta) = \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \varepsilon_F^2(t,\theta). \tag{5.15}$$

$\hat{\theta}_N$ denotes the estimate of model parameters from N data points and represents the best possible choice of parameters from the considered set $D_{\mathcal{M}}$ in the minimization of $V_N(\theta)$ function point of view. The final estimated model is

$$y(t) = G(q, \hat{\theta}_N)u(t) + H(q, \hat{\theta}_N)e(t).$$

Let us now have a look at the effect of the PEM in the frequency domain. In the following, we will use a notation $E\{s(t)\}$ for the expected value of discrete stationary signal $s(t)$ and $R_s(t)$ for the covariance of $s(t)$ (Ljung [1999]). Using the property of the inverse Fourier transform

$$E\left\{s^2(t)\right\} = R_s(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_s(\omega)d\omega,$$

Φ_s denotes (power) spectra of the signal s, we are able to reformulate the problem (5.15) into the frequency domain as

$$\hat{\theta}_N = \arg\min_{\theta} E\left\{\varepsilon_F^2(t,\theta)\right\} = \arg\min_{\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{\varepsilon_F}(\omega,\theta)d\omega. \tag{5.16}$$

By substituting (5.10), (5.11) and (5.12) into (5.16), we obtain

$$\hat{\theta}_N = \arg\min_{\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|G_0(\omega) - G(\omega,\theta)|^2\Phi_u(\omega) + \sigma_e^2|H_0(\omega)|^2}{|H(\omega,\theta)|^2}|F(\omega)|^2 d\omega$$

and for the case of a fixed noise model $H(\omega,\theta) = H_*(\omega)$

$$\hat{\theta}_N = \arg\min_{\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}|G_0(\omega) - G(\omega,\theta)|^2\underbrace{\frac{\Phi_u(\omega)|F(\omega)|^2}{|H_*(\omega)|^2}}_{F'(\omega)}d\omega. \tag{5.17}$$

(5.17) clearly shows that PEM try to fit the original transfer function $G_0(q)$ in the best possible manner in sense of the two norm with respect to frequency weighting by filter $F'(\omega)$. $\Phi_u(\omega)$ depends on the input to the system, whilst $H_*(\omega)$ on the noise expectations. Hence, the user defined filter $F(\omega)$ is the only way, how to affect the identified model. Moreover, following proposition extends the understanding to this filtering.

Proposition 5.1: *If the predictor (5.3) is time-invariant and linear in parameters and $u(t)$, $y(t)$ are scalars (implying a Single-Input Single-Output (SISO) system), then the result of filtering of one step ahead prediction error ε is the same as filtering the input-output data first and then applying the predictor (Ljung [1999]).*

This proposition will be used further in the thesis as a starting point for the incorporation of wavelets into the system identification.

6 Single LTI system identification with wavelets

6.1 Incorporation of wavelets into the system identification

The linear system identification is well established and described field. The most of the identification methods for linear systems operates in time-domain, i.e. they try to match the time-domain data as much as possible and do not take the information in the frequency domain into account. The consequence is that they try to match the whole frequency range of the system to be identified (or, more precisely, the frequency range covered by the data). In sharp contrast, the methods operating in frequency domain possess an advantage of a specific frequency range selection and are used with a benefit as an application oriented approach. A typical problem is a case of singularly perturbed system, where usually only the subsystem with slow or fast dynamics is of interest Kokotovic and Khalil [1986].

It was already shown in Chapter 4, that the DWT can be understood as a frequency filtering. Moreover, the proposition 5.1 gives the conditions for the filtering to be used directly on input-output data while also directly affecting the prediction error. The proposition however required the predictor to be linear in parameters, therefore a well-known ARX structure (Section 5.2) has been chosen. Based on (5.5), (5.7) and (5.9), the predictor for ARX model structure is of the form

$$\begin{aligned}\hat{y}(t|t-1,\theta) &= \left[1 - H^{-1}(q,\theta)\right] y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) \\ &= \left[1 - A(q,\theta)\right] y(t) + A(q,\theta)\frac{B(q,\theta)}{A(q,\theta)}u(t) \\ &= \left[1 - A(q,\theta)\right] y(t) + B(q,\theta)u(t) \\ &= -\sum_{k=1}^{n_a} a_k q^{-k} y(t) + \sum_{k=0}^{n_b} b_k q^{-k} u(t) \end{aligned} \tag{6.1}$$

and is linear in unknown parameters $a_k,\ b_k$. This equation can be rewritten as

$$\hat{y}(t|t-1,\theta) = z^T(t)\theta,$$

where $z(t) = [-y(t-1), \ldots, -y(t-n_a), u(t), \ldots, u(t-n_b)]^T$ is the vector of data measured up to time t and $\theta = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}]^T$ is the vector of unknown parameters. Expressing this equation for all measured data (more time points) and concatenating them, we obtain the set of equations $\hat{Y}(t,\theta) = Z(t)\theta$ with

$$\hat{Y}(t,\theta) = \begin{bmatrix} \hat{y}(t|t-1,\theta) \\ \hat{y}(t+1|t,\theta) \\ \vdots \\ \hat{y}(t+N|t+N-1,\theta) \end{bmatrix},$$

$$Z(t) = \begin{bmatrix} -y(t-1) & \cdots & -y(t-n_a) & u(t) & \cdots & u(t-n_b) \\ -y(t) & \cdots & -y(t-n_a+1) & u(t+1) & \cdots & u(t-n_b+1) \\ \vdots & & \vdots & \vdots & & \vdots \\ -y(t-1+N) & \cdots & -y(t-n_a+N) & u(t+N) & \cdots & u(t-n_b+N) \end{bmatrix}.$$

The solution $\hat{\theta}_N$ of the equation $\hat{Y}(t,\theta) = Z(t)\theta$ can be easily obtained as a solution to an Ordinary Least Squares (OLS) optimization problem

$$\begin{aligned} \hat{\theta}_N &= \arg\min_\theta \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left[Y(t) - \hat{Y}(t,\theta) \right]^2 \\ &= \arg\min_\theta \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left[Y(t) - Z(t)\theta \right]^2 \qquad (6.2) \\ &= \left[Z^T(t) Z(t) \right]^{-1} Z^T(t) Y(t) \qquad (6.3) \end{aligned}$$

derived within Section 5.3. $Y(t)$ is vector of measured outputs compound in the same way as $\hat{Y}(t,\theta)$, see e.g. Ljung [1999]. Note that a square of a vector $z \in \ell^2(\mathbb{Z}_N)$ is defined as $z^2 = z^T z$. For the simple notation, we will use Y, Z instead of $Y(t)$, $Z(t)$ further on.

Recall now that wavelet coefficients are to be evaluated as an inner product of the time signal and even shifts of the wavelet filters. On the $\ell^2(\mathbb{Z}_N)$ space, the inner product can be written as a vector multiplication. If z and $R_k\varphi$ from the (4.8) are vectors, then for the coefficients of the wavelet transform of the signal z the following holds

$$z * \tilde{\varphi} = \langle z, R_k\varphi \rangle = \overline{\langle R_k\varphi, z \rangle} = \overline{(R_k\varphi)^T \overline{z}} = \overline{(R_k\varphi)^T} z. \qquad (6.4)$$

For simpler notation, let us consider real valued wavelet filters, what gives us $\overline{(R_k z)^T} = (R_k z)^T$. The obtained form can be directly used for incorporation of wavelet transform into the SID problem (6.2). Note that the complex conjugate is applied on a wavelet filter. Specifically, the columns of the data matrices Y, Z represent mutually shifted input and output signals. Each of them can be transformed by wavelets according to the right side of the equation (6.4). Since the wavelet transform is linear, the problem (6.2) can be written as a multiplication with an appropriate matrix T. The problem is thus transformed onto the problem of minimizing $[TY - TZ\theta]^2$, where the

matrix $T = T(\varphi, \psi, \mathcal{P})$ contains all possible shifts of wavelet filters φ, ψ at all applied levels $\mathcal{P} \subset \{1, \ldots, p\}$ and we will call it "wavelet matrix". By adding some user defined weighting matrix W, the final optimization problem is

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left[WTY - WTZ\theta\right]^2 \tag{6.5}$$

and is still solvable via OLS. When orthogonal wavelets are used, the energy of the transformed signal is preserved in both time and frequency domains by virtue of Parseval's theorem (4.2). Hence error minimization in either domain would give the solution of the same quality Mukhopadhyay et al. [2010].

Notice that transformed prediction error $TY - TZ\theta$ can be understood in 2 possible ways, which both are mathematically identical. The first point of view is transforming data stored within Y, Z, thus analysing their time-frequency properties. The second point of view is transforming just the prediction error as $T(Y - Z\theta)$, hence analysing the time-frequency properties of the prediction error with no notion of the data themselves.

Also, at first look, there seems to be a better way of input-output data filtering (transforming), namely to transform input-output data first and then use it for creating of the matrices Y, Z. However, when transforming the data first, we obtain wavelet coefficients corresponding to wavelet filters shifted by even number of samples and they do not preserve the time-structure of the data anymore. Therefore, the consequently created matrices Y, Z would have a spoiled structure.

6.2 Wavelet matrix T

This section deals with the construction of the wavelet transform matrix T. There are two main practical limitations that have to be considered while implementing the DWT:

1. **Analysed data length:** For the purposes of a wavelet analysis it would be convenient to have data of the length of 2^p, where p is maximum analysis level. It is however obvious that it is highly impractical to guarantee this requirement.

2. **Data periodicity:** In the DWT, the periodical extension of a vector is used for analysis of the whole data. However, due to a number of reasons such as adding high frequencies due to "non-continuity" of periodically extended data or disabling of recursive identification, it is not convenient to periodically extend measured data. Therefore we will use non-extended data for wavelet identification which brings other theoretical modifications.

Let us discuss now the above mentioned limitations in detail. Let the measured data be of length D and n_a be the order of an estimated model. From the structure of Z it turns out that the length of input-output data for analysis is equal to the dimension of the column space of Z, that is $N = D - n_a + 1$, what is desired to be the power of 2. There are two possible points of view on the basic principle of wavelet analysis:

1. Both the approximations and the details of the analysed data are kept of length N (by making use of the upsampling operator) and scaled wavelet filters

$$\varphi_j = 2^{\frac{j}{2}}\varphi_1(2^j t),\ \psi_j = 2^{\frac{j}{2}}\psi_1(2^j t),\ t \in \mathbb{N}$$

   are used. Then the length of wavelet filters at the j^{th} level is $L_j = 2^{j-1}L_1$ and the shift at this level is $s_j = 2^{j-1}s_1 = 2^j$.

2. On the other hand, the lengths of both approximations and details at all levels are decreased and the analysis is always performed by the same filters φ_1, ψ_1. Consequently, the length at each level is L and shift S (generally $S = s_1 = 2$).

Since the analysed data need not be of length 2^p for any $p \in \mathbb{N}$, the latter approach is more accurate and more convenient for further use. Based on that, data of length N is decomposed (thanks to the even shifts and periodically extended vectors) into the approximations and the details, both of length $N/2$. As we do not assume periodically extended vectors, the dimension of the subspaces will be in our applications always less than $N/2$. The dimension depends not only on the data length but on the length of basic wavelet filter L as well, since $N \geq L + 2k$ has to hold, where $k \in \mathbb{N}_0$ is here maximum possible number of the shifts of the wavelet filter of the length L at the particular analysis level.

Under these assumptions, the data of length $N_1 = N$ is decomposed into the approximations and the details, both of length $N_2 = 1 + \left\lfloor \frac{N_1 - L}{S} \right\rfloor$, where $\lfloor z \rfloor$ denotes the integer part of z. This formula can be written recursively as

$$N_{j+1} = 1 + \left\lfloor \frac{N_j - L}{S} \right\rfloor \tag{6.6}$$

as long as the data are long enough for analysis at the next level. Then the number of iterations is maximum level p of wavelet analysis. With the knowledge of p and individual lengths $N_j, j = 1, \ldots, p$ the wavelet matrix T can be computed successively:

1. **At the 1^{st} level** the analysis is done by matrix T_1 which contains the even shifts of the wavelet filters in rows. Because the wavelet filter of length L could be smaller than the data length N_1, the wavelet filters φ, ψ have to be replenished onto the length N_1 with zeros for the purposes of the inner product with data. The replenished wavelet filters do not make any differences in both time and frequency domain properties and will be considered further on. Let us define

the matrix T_1 as

$$T_1 = \left[\begin{array}{c} T_{1,D}(\psi) \\ \hdashline T_{1,A}(\varphi) \end{array}\right] = \left[\begin{array}{c} \psi \\ R_2\psi \\ \vdots \\ R_{2(N_2-1)}\psi \\ \hdashline \varphi \\ R_2\varphi \\ \vdots \\ R_{2(N_2-1)}\varphi \end{array}\right], \tag{6.7}$$

where $T_{1,D}(\psi), T_{1,A}(\varphi)$ are sub-matrices of the mother wavelets and of the father wavelets, respectively. Both sub-matrices have size $N_2 \times N_1$. The transformed prediction error is now

$$T_1 Y - T_1 Z\theta = \left[\begin{array}{c} T_{1,D}(\psi)Y \\ \hdashline T_{1,A}(\varphi)Y \end{array}\right] - \left[\begin{array}{c} T_{1,D}(\psi)Z \\ \hdashline T_{1,A}(\varphi)Z \end{array}\right]\theta. \tag{6.8}$$

The upper part express a transformed set of the equations into the subspace of details and lower part into the subspace of approximations. Note that size of $T_1 Y$ is $2N_2 \times 1$ and size of $T_1 Z$ is $2N_2 \times (n_a + n_b + 1)$.

2. **Higher levels** analysis is in principle very similar to analysis at the 1^{st} level. The j^{th} level analysis, $j \in \mathbb{N}, j > 1$, could be characterized by multiplying by the matrix T_j. In Chapter 4 it was said that the j^{th} level analysis of measured data is the same as the 1^{st} level analysis of the approximation of measured data at the $(j-1)^{th}$ level. Details at all levels up to the $(j-1)^{th}$ level have to be preserved without change.

   Let us treat the matrix T_2 for the 2^{nd} level first and higher levels will be shown analogously afterwards. Based on a similar idea as before with matrix T_1, the matrix T_2 is defined as

$$T_2 = \left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & \overline{T_2} \end{array}\right], \text{ where } \overline{T_2} = \left[\begin{array}{c} T_{2,D}(\psi) \\ \hdashline T_{2,A}(\varphi) \end{array}\right]. \tag{6.9}$$

   The unitary $N_2 \times N_2$ matrix I serves for the preservation of the details at the 1^{st} level. The matrix $\overline{T_2}$, defined analogously to the matrix T_1, comprises the wavelet filters replenished onto the length of N_2. Consequently the size of $\overline{T_2}$ is $2N_3 \times N_2$ and the total size of the matrix T_2 is $(N_2 + 2N_3) \times 2N_2$. The whole wavelet matrix for transformation at the 2^{nd} level is then $T = T_2 T_1$.

Finally, T_j can be expressed as follows:

$$T_j = \left[\begin{array}{c:c} I & 0 \\ 0 & \overline{T_j} \end{array}\right], \text{ where } \overline{T_j} = \left[\begin{array}{c} T_{j,D}(\psi) \\ \hdashline T_{j,A}(\varphi) \end{array}\right] = \left[\begin{array}{c} \psi \\ R_2\psi \\ \vdots \\ R_{2(N_{j+1}-1)}\psi \\ \hdashline \varphi \\ R_2\varphi \\ \vdots \\ R_{2(N_{j+1}-1)}\varphi \end{array}\right]. \quad (6.10)$$

It is obvious that the wavelet filters have to be replenished onto the length of N_j. The unit matrix I has now the size $\sum_{i=2}^{j} N_i \times \sum_{i=2}^{j} N_i$, matrix $\overline{T_j}$ has size $2N_{j+1} \times N_j$ and finally, the whole wavelet transform matrix is

$$T = \prod_{j=1}^{p} T_j = T_p \cdot \cdots \cdot T_1.$$

So far, we have dealt with the construction of the matrix T from the specific (selected) wavelet basis function only. There is, however, a possibility to exploit several wavelet functions within the matrix T at once. Realizing that the matrix T contains one row (wavelet function) for each specific (and possible) shift and scale of the wavelet φ or ψ, the main reason for utilizing of more than one basic wavelet function is obviously to increase the number of equation comprising the information from the particular time and frequency localisation.

There arises yet another question: Why to use several banks of wavelet functions if every wavelet function divides the frequency domain in the same way (Section 4.4)? The answer is that there are differences between those functions. The Figure 6.1 depicts two wavelets db4 and db8, both at the 5^{th} analysis level, in both time and frequency. The first difference is the support of those functions in the time domain (as already commented in Section 4.5.1) and the second difference is the steepness of their frequency characteristics. The larger the support is, the larger part of data is investigated. The steeper characteristic of the filter is, the more precise is the filtering.

6.3 Weighting matrix W

The weighting matrix W is a user-defined, diagonal matrix with its elements as weighting coefficients for the particular wavelet filters given by their shift and scale. This section discusses several possibilities for choice of the weighting matrix.

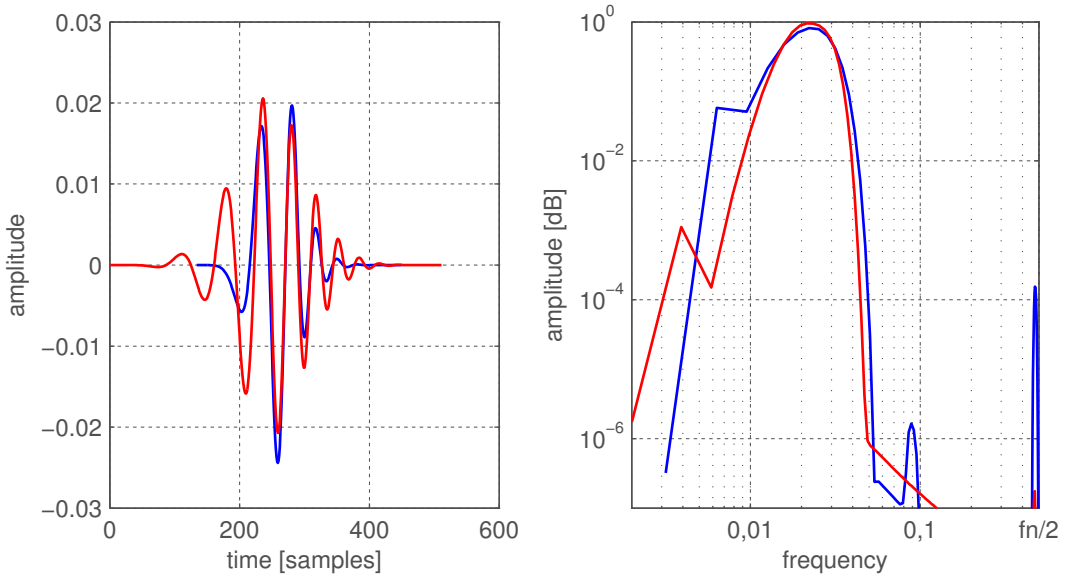


Figure 6.1: Difference between distinct wavelet filters at the same level.

A quite well-known way of selecting of particular weights is to use the power spectra or empirical transfer function estimate computed from the input-output data Ljung [1999]. This method computes the weights for a particular frequency range in accordance to the energy present in that range, thus the weights are assigned in accordance to the shape of the frequency characteristics. Such a procedure seems to be something like automatic computation of weights, however there are two drawbacks: i) The weights need not correspond to the application and to the demands on the model (they only correspond to the information in the data). ii) Small signal to noise ration can break the spectra on higher frequencies, hence can enlarge particular weights and consequently destroy the resulting model.

The previous method weights each frequency range by the same weight without considering the time domain. However, data can contain some time intervals where the specific dynamics can be observed, thus it is worth to put an emphasis on these intervals by weighting of appropriate wavelet functions, which are localised in that time interval. The procedure can be for example done in two steps: i) One weighting matrix contains only ones and zeros, what causes the selection of appropriate wavelet filters. ii) Another weighting matrix contains weights of those selected filters to evaluate their significance. A typical situation for this procedure is when the model is identified from the data obtained during an experiment[1] on the system. Such a data-set can contain (and usually contains in practise) specific time intervals, where particular property of the system is tested, so only some appropriate wavelet filters are worth to use.

[1] To obtain the sufficiently excited data containing sufficient information about the system, a so called "experiment" is performed on the system, where the specific, pre-computed input is applied.

Another possibility of weighting is to adapt the weights on particular frequency range to the number of wavelet filters covering this range. Since wavelet filters on higher levels (scales) have longer support and since the length of data is given, the number of possible shifts decrease with each next level. Then, the sum of square of prediction error corresponding to the particular frequency range becomes smaller only due to the smaller number of equations. A remedy can be to set weights inversely proportional to the number of wavelet filters on the particular frequency range.

Next, due to partial overlapping of wavelet filters in frequency domain (see e.g. Figure 4.2), the filters from table Table 4.1 (made by composition of basic wavelets) have not unit gain at any frequency. This effect (together with constant coefficients belonging to those filters) can be compensated by normalizing the wavelets by weighting matrix W computed from the vector of weights

$$V = \left[\frac{1}{\max_\omega(\hat{w}_1(\omega))}, \ldots, \frac{1}{\max_\omega(\hat{w}_p(\omega))}, \frac{1}{\max_\omega(\hat{w}_{p+1}(\omega))} \right], \tag{6.11}$$

where $\hat{w}_j,\ j = 1, \ldots, p+1$ are filters from the table Table 4.1, $j = 1, \ldots, p$ for details at the j^{th} level and $j = p+1$ for approximation at the p^{th} level.

Yet another constraint can be applied to the matrix W. It has already been mentioned that filters overlap each other in frequency domain. Due to this fact and user-friendly use, weights of filters should be chosen such that no filter overweights any other, especially the adjacent filters. In consequence, there is some upper bound (or lower bound, it depends on point of view) for weights for each particular wavelet filters families Heil and Walnut [1989]; Mallat [1999]; Rao et al. [1999].

Finally, the total weighting matrix W can be computed from individual weighting matrices $W_1, \cdots, W_q,\ q \in \mathbb{N}$ as

$$W = \prod_{j=1}^{q} W_j = W_q \cdot \cdots \cdot W_1.$$

6.4 Asymptotic properties of the estimate

We have shown how to incorporate the wavelet transform into model identification. Now, let us have a look at the limit behavior of the algorithm, which is very important for practical use. It was shown in Ljung [1999] that the ARX model is strictly globally identifiable. Let θ^* denote such a model parameters vector, which is the best possible theoretical solution of the (6.2) (when $N \to \infty$), and let $\hat{\theta}_N$ denote the solution computed from measured data (some finite N). The convergence of the estimate of the ARX model parameters via wavelets and its quality are investigated:

1. **Convergence:** For general PEM it is proven that if noise $e(t)$ is not correlated with regressor Z, the following holds:

$$\lim_{N\to\infty} \hat{\theta}_N = \theta^*. \tag{6.12}$$

This condition is not satisfied in the case of the ARX model and the estimate is biased.

$$\lim_{N\to\infty} \hat{\theta}_N = \theta^* + \lim_{N\to\infty} E\left\{(Z^T Z)^{-1} Z^T e\right\}. \tag{6.13}$$

Since the proposition 5.1 was complied, this limit has to hold analogously with using of wavelets for filtering:

$$\lim_{N\to\infty} \hat{\theta}_N = \theta^* + \lim_{N\to\infty} E\left\{(Z^T T^T W^2 T Z)^{-1} (WTZ)^T e\right\}. \tag{6.14}$$

Note that the fact that the estimate is biased does not mean that the estimate is wrong. The bias is caused by the demand on parameters estimate to give the best possible predictor (Ljung [1999]). It consequently means that we really obtain the best possible predictor, what need not be the best approximation of the true system from the simulation point of view.

Moreover, there is the instrumental-variable (IV) method (for details see e.g. Van den Hof [1996]) which guides the user how to have uncorrelated data Z with noise e, hence to have an unbiased estimate. Also realize that when it is possible to have unbiased estimate just by adjustment of the matrix Z, then the best predictor depends on the measured data, what consequently leads to the condition of sufficient excitation.

2. **Quality of the estimate:** The estimation error $G_0(q) - G(q, \hat{\theta}_N) = G_0(q) - G(q, \theta^*) + G(q, \theta^*) - G(q, \hat{\theta}_N)$. The term $G_0(q) - G(q, \theta^*)$ is a structural error. It describes an incapability of the model (with given structure) to match the system. The meaning could be well understood in frequency domain. If the fixed noise model is considered, the structural error is minimum if and only if the solution θ^* is found by solving least squares problem $\min_\theta |G_0(\omega) - G(\omega, \theta)|^2$. The term $G(q, \theta^*) - G(q, \hat{\theta}_N)$ is an error caused by noise affecting the measured data and represents how far from the optimal solution the found solution is. This can also be comprehended as variation of the random variable $\sqrt{N}\left(\hat{\theta}_N - \theta^*\right)$. In case of an open-loop, the variance of the frequency function estimate at certain frequency ω can be written as (Zhu [2001])

$$\mathrm{Var}G(\omega, \hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}, \tag{6.15}$$

$v(t) = H(q, \hat{\theta}_N) e(t)$ is filtered noise. For further derivation, let us remind, that (power) spectra of the response $y(t) = G(z)u(t)$ of the system $G(z)$ on the input $u(t)$ is $\Phi_y(\omega) = \Phi_u(\omega)|G(e^{i\omega})|^2$. Obviously, each wavelet filter can be viewed as a system with particular frequency characteristic $\hat{w}_j(\omega)$. Because each of these systems filters the same input data and because of filters (practical) exclusive frequency localisation, the total variance satisfies

$$\left(\mathrm{Var}G(\omega, \hat{\theta}_N)\right)^{-1} = \sum_{j=1}^{p+1} \left(\mathrm{Var}_j G(\omega, \hat{\theta}_N)\right)^{-1}, \tag{6.16}$$

$\mathrm{Var}_j G(\omega, \hat{\theta}_N)$ stands for variance caused by wavelet filter $\hat{w}_j(\omega)$. This partial variance can be written as

$$\mathrm{Var}_j G(\omega, \hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega) \left|V'(j)\hat{w}_j(\omega)\right|^2}, \tag{6.17}$$

where

$$V'(j) = \frac{V(j)}{\max\limits_k \{V(k)\}} \tag{6.18}$$

is the normalized weight for the j^{th} level analysis, $V(k)$ is k^{th} element of the vector of weights from which the matrix W is constructed. The normalized weights have to be considered to correctly derive the variance $\mathrm{Var}_j G(\omega, \hat{\theta}_N)$. Let us make the substitution $\Gamma_j = \left|V'(j)\hat{w}_j(\omega)\right|^2$. Then

$$\left(\mathrm{Var} G(\omega, \hat{\theta}_N)\right)^{-1} \approx \sum_{j=1}^{p+1} \left(\frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)\Gamma_j} \right)^{-1} = \frac{N}{n} \frac{\Phi_u(\omega)}{\Phi_v(\omega)} \sum_{j=1}^{p+1} \Gamma_j \tag{6.19}$$

and the total variance

$$\mathrm{Var} G(\omega, \hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \left(\sum_{j=1}^{p+1} \left|V'(j)\hat{w}_j(\omega)\right|^2 \right)^{-1}. \tag{6.20}$$

In case of no additional weighting, the sum equals 1 and the results of the identification with and without wavelet filtering are very close to each other[1] both in time and frequency domain. On the other hand, additional weighting affects the variance of the estimate at the particular frequency range and, consequently, the behavior of the estimated model in this frequency range as well. And as was already mentioned, this change of variance of the estimate behaves in order to the Parseval's theorem (4.2), thus in order to the coefficients of wavelet transform with appropriate wavelets.

To show a parallel to understanding the problem of the ARX model parameters estimation in the frequency domain, the following formulae evaluating the quality of the estimated parameters holds:

$$\mathrm{trace}\{\mathrm{cov}\{\theta^* - \hat{\theta}_N\}\} = \sigma_e^2 \mathrm{trace}\left\{ \left(Z^T T^T W^2 T Z \right)^{-1} \right\}.$$

6.5 Possible extensions

This section shows possible improvements of the estimates useful especially in applications.

[1] Strictly speaking, variances of the estimate $G(\omega, \hat{\theta}_N)$ would be the same, the estimate itself however need not be.

1. **Thresholding of wavelet analysis.** The thresholding in the wavelet analysis is used to reduce noise. This operation means to nullify such the wavelet coefficients which are lower than some threshold $\varepsilon_t \in \mathbb{R}$. Consequently, the threshold provides the lower limit for the considered portion of particular frequency range in the original signal. Globally, it can lead to more accurate numerical results but, on the other hand, some information from the input-output data is removed.

   Let $\eta_y,\ \eta_z > 0$ are values of thresholds for certain wavelet type ψ, certain shift $t \in \mathbb{N}$ and certain analysis level j. Next, denote $y_c,\ z_c$ some columns of matrices $Y,\ Z$, respectively. Then if any of wavelet coefficients are lower than the appropriate thresholds:

   $$y_c * \tilde{\psi}_j(t) < \eta_y \ ,\ z_c * \tilde{\psi}_j(t) < \eta_z,$$

   the appropriate row (corresponding to some equation) within the set $TY - TZ\theta$ is nullified (omitted). This procedure is performed during the analysis for all wavelet filters, each analysis level can have (and usually have) different thresholds in order to required amount of information in corresponding frequency band.

2. **Keeping of the wavelet analysis coefficients at lower levels:** As mentioned before, the approximation at the j^{th} level is replaced by the approximation and the details, both at the $(j+1)^{th}$ level. The consequence is that still less and less equations are generated for the approximation, thus for a low-frequency content of data. Moreover, since low-frequency wavelets have large support and larger basic shift than high-frequency wavelets, it is natural that there are much less low-frequency equations. As a remedy, we can double the number of equations on each level and then keep one half and use the second half for the analysis at the next level. The procedure is still done in a sense of (6.5), but with

   $$T_j = \left[\begin{array}{c:c} I & 0 \\ 0 & I \\ \hdashline 0 & \overline{T_j} \end{array}\right],$$

   $\overline{T_j}$ remains the same as in (6.10). This significantly expands the number of equations, thus it can improve quality of the result. However it is recommended to apply it only for frequency ranges of interest because of increases in computational demands.

3. **Recursive identification:** Recursive solution of the OLS problem (6.2) is described in e.g. Engel et al. [2004]; Ljung and Ljung [1985]. Recursive identification applied to SID using wavelet transform is more complicated. There is one inherent difference from a classical recursive identification, namely, a length of newly measured data has to be, at least, the length of shift of any wavelet filter used for the SID. Hence the minimum length of new data is shift of those used wavelets which cover the highest frequencies of interest, and on the contrary, for one new equation in the lowest frequencies of interest, the length of new data has to be at least the shift of the appropriate (maximum level used) wavelet filters. (Note

that the maximum level used need not be the maximum possible level, since the choice of used levels depends on the application while the maximum possible level is given by the length of the data. Such the situation turns up in cases of identifying a faster system's dynamics while the information about slow dynamics is not of interest. Moreover, the weighting matrix W could also suppress particular frequency ranges.) This condition must be considered due to possibility to perform at least 1 new shift of any wavelet filter whereby at least 1 new coefficient (equation) is obtained. If this condition is satisfied, the prediction error can be written as

$$\begin{bmatrix} T_{old} \\ T_{new} \end{bmatrix} \begin{bmatrix} Y_{old}(t) \\ Y_{new}(t) \end{bmatrix} - \begin{bmatrix} T_{old} \\ T_{new} \end{bmatrix} \begin{bmatrix} Z_{old}(t) \\ Z_{new}(t) \end{bmatrix} \theta, \tag{6.21}$$

where $T_{old} = \begin{bmatrix} T & 0 \end{bmatrix}$, $Y_{old} = Y$, $Z_{old} = Z$. The construction of Y_{new}, Z_{new} is based on the new input-output data and T_{new} is matrix of all new possible shifts of wavelet filters. After multiplication

$$\begin{bmatrix} TY(t) \\ T_{new} \begin{bmatrix} Y(t) \\ Y_{new}(t) \end{bmatrix} \end{bmatrix} - \begin{bmatrix} T(t) \\ T_{new} \begin{bmatrix} Z(t) \\ Z_{new}(t) \end{bmatrix} \end{bmatrix} \theta, \tag{6.22}$$

which leads to the expanded predictor

$$TY(t) - TZ(t)\theta \tag{6.23}$$

$$T_{new} \begin{bmatrix} Y(t) \\ Y_{new}(t) \end{bmatrix} - T_{new} \begin{bmatrix} Z(t) \\ Z_{new}(t) \end{bmatrix} \theta \tag{6.24}$$

Weighting matrix W is also expanded in an appropriate manner. Unfortunately, the newly arisen set of equations (6.24) depends on the whole set of the input-output data contained in $Z(t), Y(t)$, thus for parameter θ actualization is necessary to recompute it again from the whole data set. Any only advantage is provided if maximum analysis level used is lower than maximum theoretical analysis level. In that case, an "unused" part of the matrix T_{new} is filled by zeros ($T_{new} = \begin{bmatrix} 0 & T'_{new} \end{bmatrix}$) what means less input-output data is needed to be stored in memory for Y, Z in equation (6.24). However, due to necessity of using quite large part of old data, WT is not suitable to be used within recursive identification.

6.5.1 Discussion on other model structures

Talking about the recursive identification, we should also mention the applicability of the presented method to other model structures. So far, we have dealt with ARX model structure because of the reason that the predictor is linear in parameters. There are actually several different SISO model structures which are also quite often used in practise: ARMAX and OE model structures. The reader can find more details about them for example in Ljung [1999]; Van den Hof [1996]; Zhu [2001]. These structures

enable flexible description of disturbance part of a model, what consequently yields the regressor $z(t)$ contains the error of the estimate, thus depends on the parameters. This is simultaneously their main disadvantage, since the predictor is then of the form $\hat{y}(t) = z(t,\theta)\theta$, what is called pseudolinear regression and what is impossible to solve quickly by OLS. The solution to this identification problem is usually computed using recursive least squares and as was previously mentioned, the recursive identification is not advantageous when using with wavelets. Nevertheless, when transforming the regressor in the presented way, one can utilize a recursive identification without any change, the features of incorporation wavelets into SID remain the same. All adaptations of the recursive algorithm like forgetting can also be applied.

6.6 Case study

The proposed algorithm was implemented and tested on an example thereby also demonstrating the main features of the algorithm. Only a simulation example is used for the reason of showing how the proposed algorithm performs, real-life cases will be treated within the next section dealing with multivariable system identification. Consider simple system with transfer function

$$G = \frac{0.0012272(s+178)(s+2)^3}{(s+0.1)^2(s^2+4.8s+174.8)}.$$

The objective of this example is to identify two models, each of the 2^{nd} order and each describing slow or fast dynamics of the original system, respectively. The frequency characteristic of the system is well divisible into slow and fast part - poles are $\{-0.1\ ,\ -0.1\}$ and $\{-2.4-13i\ ,\ -2.4+13i\}$ for slow and fast dynamics, respectively. This example is therefore very suitable for the demonstration of how wavelet filters can work within SID. For simulation purposes, the system was discretized by sampling time $T_s = 0.1$ s, thereby we obtained

$$G_d = \frac{0.0012272(z+13.94)(z-0.6902)(z^2-1.722z+0.7431)}{(z-0.99)^2(z^2-0.4208z+0.6188)}.$$

Let us call the ARX model obtained via wavelets as "WAV" model. The results of both the ARX and the WAV model parameter identification with all filters' weights set to 1 are depicted in Figure 6.2 and in Figure 6.3. Daubechies wavelets db4 were used and the maximum possible analysis level $p = 8$ was used. Both models are not satisfactory at all since they try to describe the global behavior of the original system, but both are of the insufficient order. Despite the fact that the system has been excited sufficiently at all necessary frequencies (up to half of the Nyquist frequency), this case also illustrates that the WAV model implicitly contains more information about the highest frequencies. It is caused by the fact that in the higher frequencies, there are more equations in the predictor, therefore more emphasis is put on higher frequencies. Note that when the order of the identified model would be 4 as of the original system,

both models would be identified exactly even without tuning the weights, since the original system G_d belongs to the set of all models with ARX structure of the 4^{th} order (the set $\mathcal{M}$), since this model structure is globally identifiable (Ljung [1999]) and since the system was sufficiently excited by data.

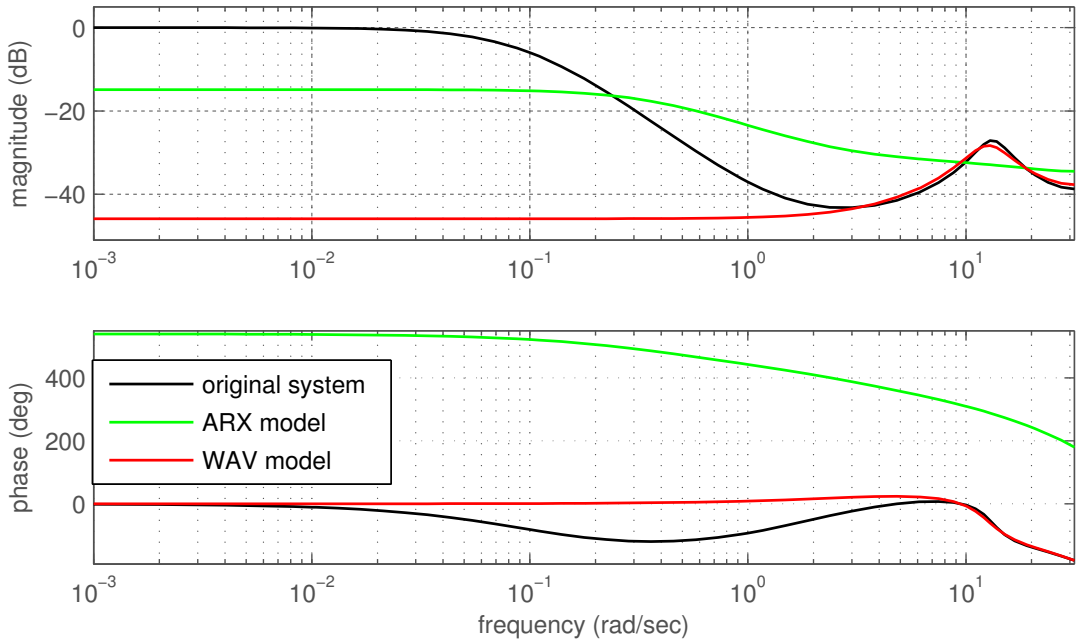


Figure 6.2: Model identified on full frequency range - Bode plot.

Table 6.1: Table of identified models.

Frequencies	2^{nd} **order model identified by**	
	ARX	WAV
All	$\frac{-0.0001038z^2+0.01856z-0.02404}{z^2-1.117z+0.1484}$	$\frac{0.00004624z^2+0.01606z-0.01076}{z^2-0.5007z+0.5538}$
Low	$\frac{0.008883z^2-0.01642z+0.007671}{z^2-1.972z+0.9717}$	$\frac{0.0108z^2-0.02032z+0.009644}{z^2-1.975z+0.9755}$
High	$\frac{0.0003624z^2+0.01583z-0.007988}{z^2-0.399z+0.6096}$	$\frac{-0.00009172z^2+0.01579z-0.007404}{z^2-0.3879z+0.6149}$

During the identification procedure, the Empirical Transfer Function Estimate (ETFE) and estimation of the frequency response together with wavelet filters (Figure 6.4) is displayed for the user convenience to adjust the filters' weights in accordance to his/her knowledge of the system. Concerning the selection of the weights of

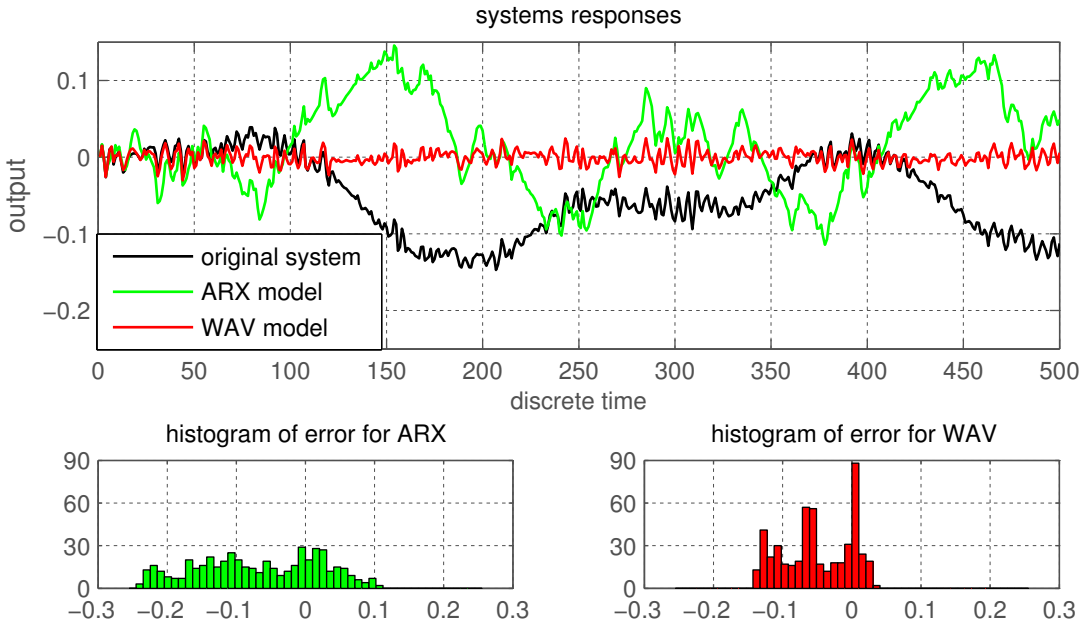


Figure 6.3: Model identified on full frequency range - simulation and histograms of errors.

the wavelet filters for slow and fast models in this case study, only the compensation of the effect of the wavelet filters overlapping was considered. To obtain credible results of the WAV model it is suitable to compare it with the ARX model pre-filtered with classical filter F. The filters were chosen as $F_{low} = \frac{1}{(s+2)^4}$ for low frequencies and $F_{high} = \frac{s^4}{(s+2)^4}$ for high frequencies. Both filters were discretized by T_s and both are depicted in bode plots. The results are depicted in figures 6.6 and 6.7 and in figures 6.8 and 6.9 for slow and fast model identification, respectively. Both bode plots clearly show that the frequency range selected for identification is perfectly covered by the models obtained by both methods. For the slow dynamics (frequencies up to 1 rad·s^{-1}, see Figure 6.4), 4 wavelet filters with the smallest frequencies were used ($\hat{w}_i,\ i = 6, \ldots, 9$ from Table 4.1 for $p = 8$) and for the fast dynamics (frequencies from 1 rad·s^{-1}), filters $\hat{w}_i,\ i = 1, \ldots, 5$ were used. The behavior of the models is worthy to verify in the time domain as well. The time responses of the slow models match the trend and higher frequencies are involved in the simulation error depicted by histograms. On the other hand, the fast models cover only fast and short-time changes and the whole trend of the signal is omitted. Numerical results are summarized in Table 6.1.

Let us now discuss the resulting ARX and WAV models. They are very similar one to each other even from the numerical point of view. The question arises why to pre-filter the identification data by wavelet filters. Wavelet filters have big advantages in comparison to the filters designed in a classical way. Firstly, they have a simple structure in the frequency domain, secondly, they complement each other in the frequency

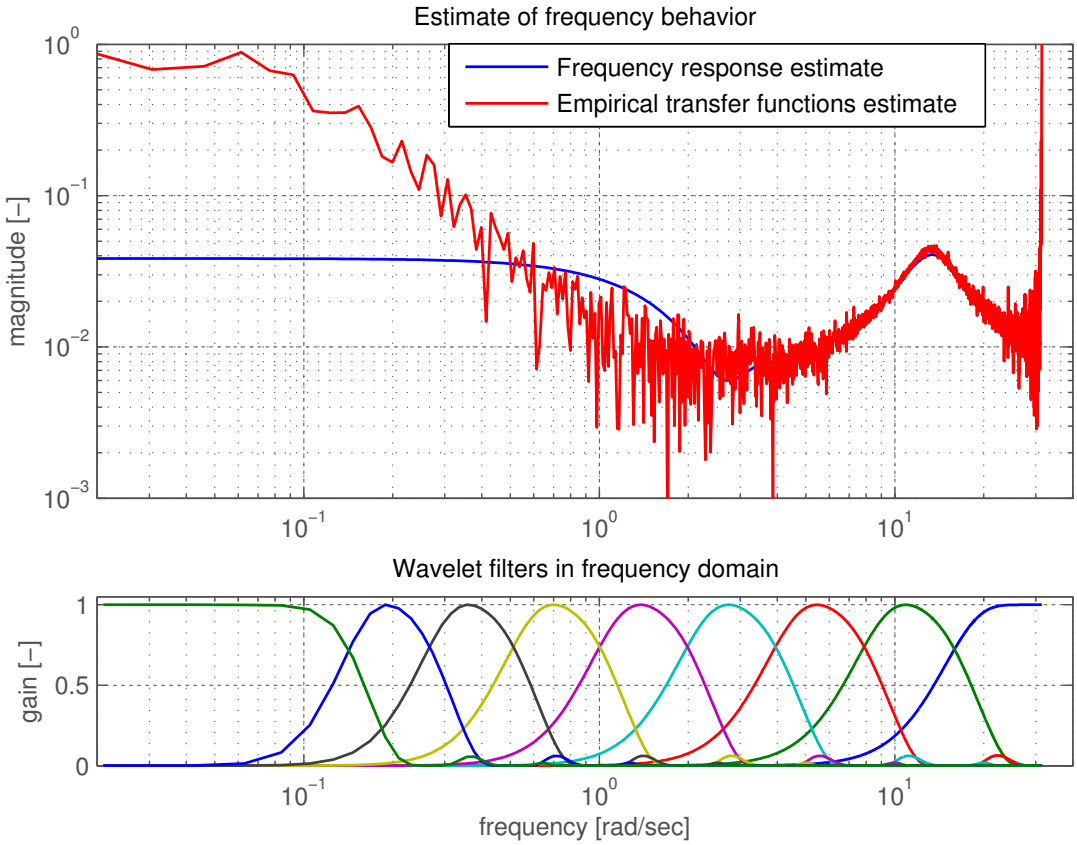


Figure 6.4: Filters and ETFE for filters weights selection.

domain. This provides us with a big advantage in real problems, where the frequency characteristics of the system to be identified are not known a priori. The satisfactory results can be acquired by tuning of the weights (which corresponds to some knowledge of the system) only. Note that once implemented, this method is quite generic, while the design of the appropriate filters can be quite time consuming (compare the simplicity of selection of weights when looking at the Figure 6.4 against describing a suitable filter of some complex shape in frequency domain). Moreover, the wavelet filters are orthogonal in the time domain, therefore each filter extracts the specific portion of information from the signals without any duplicity. This fact also contributes to the conditionality of the identification algorithm.

Moreover, there is a difference between the proposed algorithm and classical filtering utilized above, as illustrated by Figure 6.5. The classical approach to frequency-domain system identification is to first filter the input data, apply those inputs to the system and perform identification afterwards (see Figure 6.5(a)). This approach has a major drawback – the user should have a good knowledge of the frequency properties of the system in advance to take the important frequencies into account while designing the

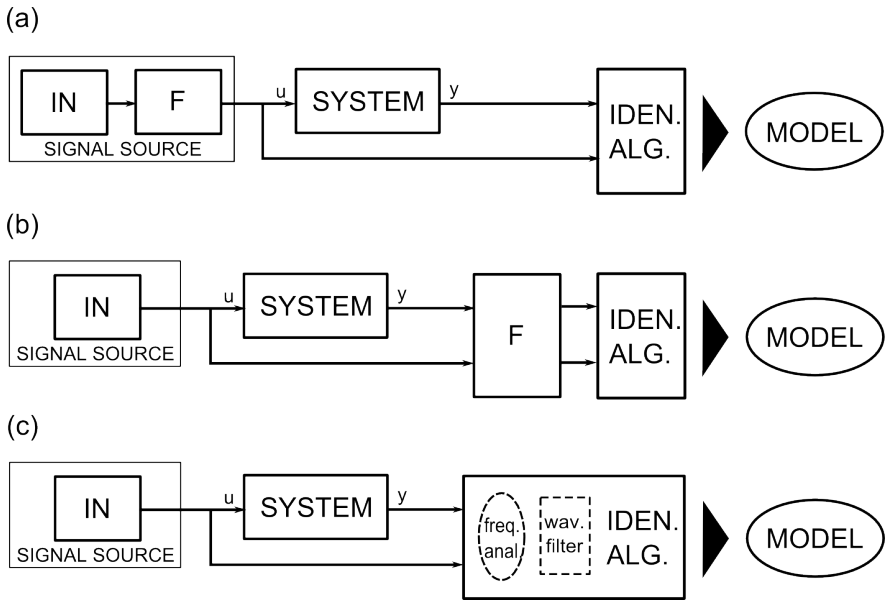


Figure 6.5: The main principle of the proposed identification algorithm. *(a,b) shows classical frequency-domain identification approaches, (c) is the modification with wavelets according to the thesis.*

filter. Moreover, if any resonance or another important frequency is not excited, the ability of the resulting model to cover this frequency decreases. The aforementioned problem can be treated to a certain extent by using filtering after applying inputs on the system (Figure 6.5(b)). Our (wavelet based) approach performs the filtering and frequency analysis on unfiltered input and output signals as well. However, to keep the time structure of the SID problem, it is impossible to separate the wavelet transform and SID. Thus, both frequency analysis and WT have to be included in the identification algorithm as illustrated in Figure 6.5(c).

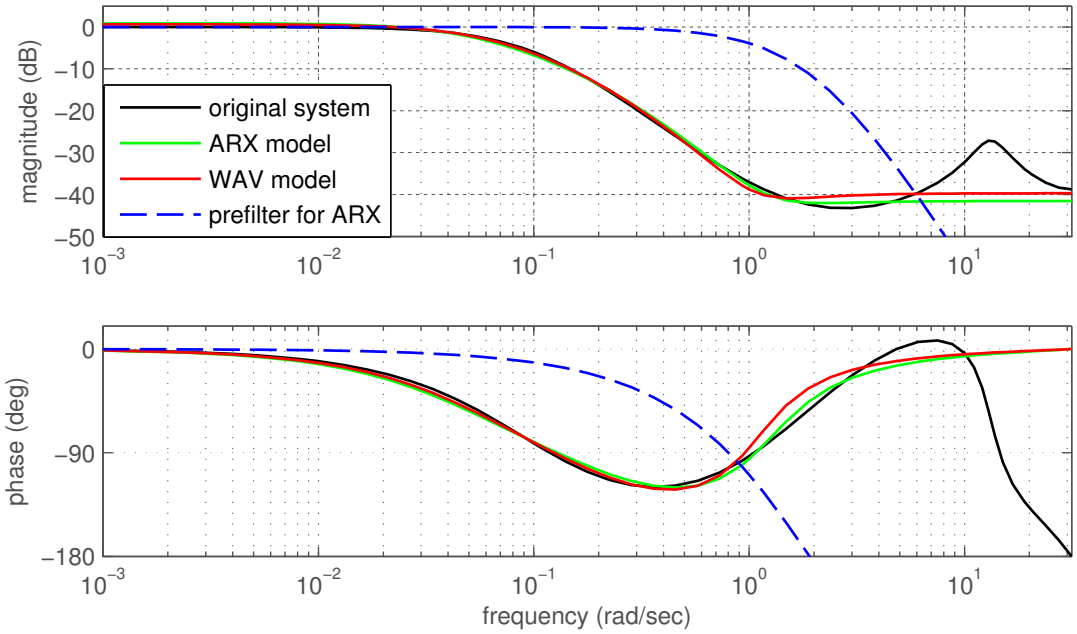


Figure 6.6: Model identified on low frequency range - Bode plot.

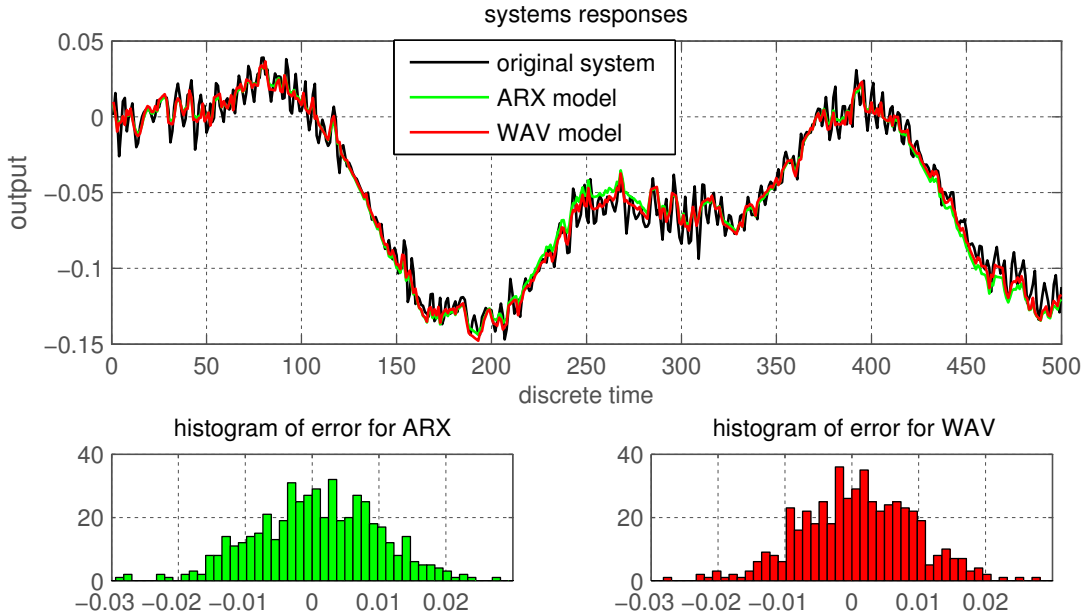


Figure 6.7: Model identified on low frequency range - simulation and histograms of errors.

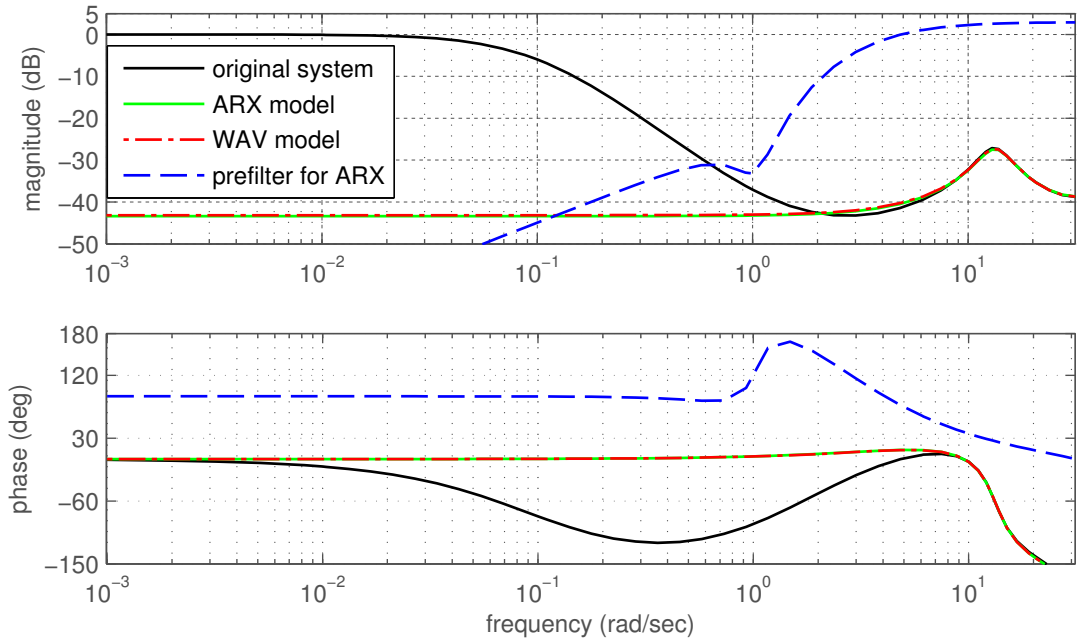


Figure 6.8: Model identified on high frequency range - Bode plot.

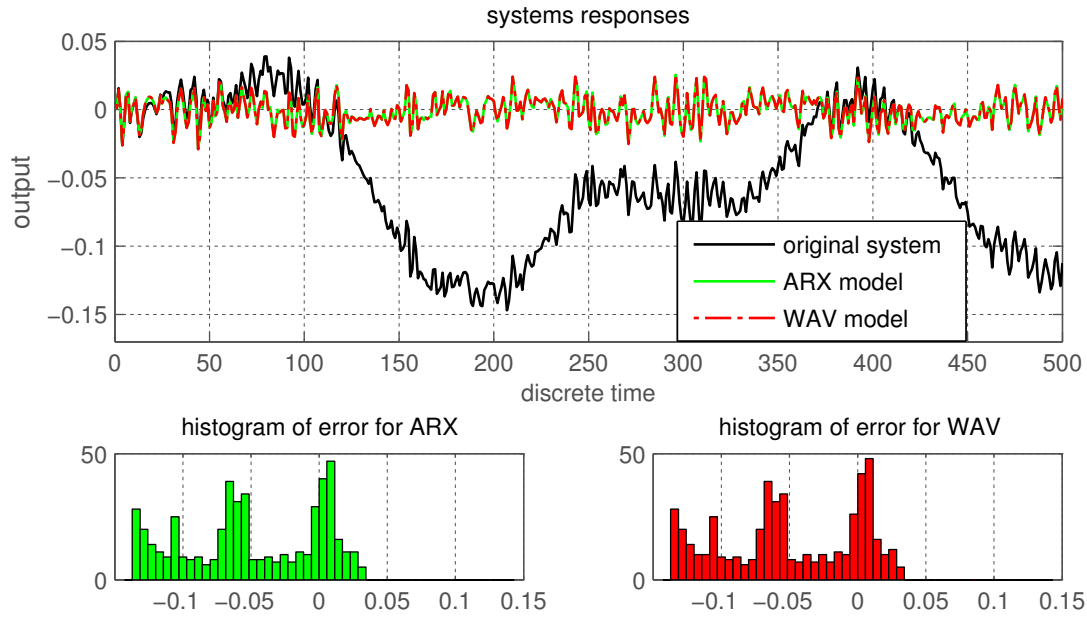


Figure 6.9: Model identified on high frequency range - simulation and histograms of errors.

7 Multivariable LTI system identification with wavelets

As the Chapter 5 introduces the system identification problem from the SISO systems point of view, the Chapter 6 shows how to exploit wavelets within the identification of a SISO system. Additionally, there are several methods to identify unknown parameters of multivariable LTI models as well. A multivariability is usually understood as that the system has more than one input or output. In order to that, the control community distinguish between two classes of multivariable systems, which are commonly marked as Multiple-Input Single-Output (MISO) systems or Multiple-Input Multiple-Output (MIMO) systems. The main SID procedure steps like choice of modeling purpose, choice of model structure and choice of estimation method remain the same as for SISO SID or at least analogical, however, the estimation method for multivariable systems can be in general more complex, more computational demanding (sometimes even intractable) and, additionally, they bring new problems into the SID theory and practise.

Although there are number of possible descriptions of LTI systems, two of them are much more frequently used then the others: a state space system description and a transfer function description. The choice of description can be largely assumed as a part of choice of appropriate model structure. There are also some typical SID methods for usual multivariable model structures: the already mentioned PEM for the transfer function description and Subspace State Space System IDentification (4SID) method (sometimes just called "subspace" method) for the system in a state space description. Moreover, these methods can be extended to be able to handle the constraints on unknown parameters Prívara et al. [2010, 2012].

In this chapter, possible incorporation of WT into the identification of multivariable systems is investigated. The Chapter 6 has revealed a condition for SID method to enable the incorporation of the DWT into it while the meaning of the DWT and its principles remain. That condition is a suitable expression of the identification problem so that a time-structure of measured data is preserved. Realize that when this structure is broken, the DWT can still be applied, but the meaning of filtering as well as time and frequency localisation are completely lost and the DWT becomes just a transformation tool with no interpretation. Therefore, in this chapter, we will address only that multivariable SID methods which are suitable for the DWT.

Typical representative of method which breaks the time-structure of data is subspace method. Subspace method exploits orthogonal projections between input and output data to compute a matrix, which is to be divided further via SVD decomposition. Both projections and SVD decomposition break the time-structure of data, thus the physical meaning of the DWT as well.

7.1 MISO LTI system identification with wavelets

It has been already mentioned that the main reason for the ARX model structure to be treated within the Chapter 6 is proposition 5.1. Nevertheless, the ARX model structure can be expanded for a MISO systems in the following way.

Let the MISO system with m inputs and 1 output be described (analogue to (5.6)) by a linear difference equation

$$y(t) + \sum_{k=1}^{n_a} a_k y(t-k) = \sum_{i=1}^{m} \sum_{k=0}^{n_b} b_{i,k} u_i(t-k) + e(t),$$

where u_i stands for the i^{th} input and $b_{i,k}$ is the coefficient b_k for the input u_i. By substitution

$$A(q,\theta) = 1 + \sum_{k=1}^{n_a} q^{-k} a_k \ , \ B_i(q,\theta) = \sum_{k=0}^{n_b} q^{-k} b_{i,k},$$

the model structure is

$$A(q,\theta)y(t) = \sum_{i=1}^{m} B_i(q,\theta)u_i(t) + e(t).$$

and transfer functions

$$G(q,\theta) = \left[\frac{B_1(q,\theta)}{A(q,\theta)}, \frac{B_2(q,\theta)}{A(q,\theta)}, \ldots, \frac{B_m(q,\theta)}{A(q,\theta)}\right] \ , \ H(q,\theta) = \frac{1}{A(q,\theta)}.$$

Afterward, the predictor is of the form

$$\begin{aligned} \hat{y}(t|t-1,\theta) &= \left[1 - H^{-1}(q,\theta)\right] y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) \\ &= \left[1 - A(q,\theta)\right] y(t) + \left[B_1(q,\theta), B_2(q,\theta), \ldots, B_m(q,\theta)\right] u(t) \\ &= -\sum_{k=1}^{n_a} a_k q^{-k} y(t) + \sum_{i=1}^{m} \sum_{k=0}^{n_b} b_{i,k} q^{-k} u_i(t) \end{aligned} \tag{7.1}$$

and is still linear in unknown parameters a_k, $b_{i,k}$. Like before, this equation can be rewritten as

$$\hat{y}(t|t-1,\theta) = z^T(t)\theta$$

with

$$\begin{aligned} z(t) &= \left[-y(t-1), \ldots, -y(t-n_a), u_1(t), \ldots, u_1(t-n_b), \ldots, u_m(t), \ldots, u_m(t-n_b)\right]^T, \\ \theta &= \left[a_1, \ldots, a_{n_a}, b_{1,0}, \ldots, b_{1,n_b}, \ldots, b_{m,0}, \ldots, b_{m,n_b}\right]^T. \end{aligned}$$

For all measured data, we again obtain the set of equations $\hat{Y}(t,\theta) = Z(t)\theta$ with $\hat{Y}(t,\theta)$, $Z(t)$ composed in a similar manner as in Section 6.1.

Since this method extends the SISO SID from the Chapter 6 and since it is actually the same class of the SID problem, its properties remain in principle the same (see Section 6.4) except for one difference. This difference is that input signals must not be mutually correlated with one another for the parameters to be strictly globally identifiable. The reason is that strongly correlated inputs are then difficult to distinguish on the output of the system. Numerically, it causes the collinearity among columns in the regressor $Z(t)$[1], hence there is some non-empty right null-space of this matrix and it consequently brings some additional degrees of freedom into the resulting set of "optimal" parameters $\hat{\theta}_N$. Afterwards, there is no change in T, W matrices and the whole principle of the DWT within SID is preserved just like in Chapter 6. Note that the collinearity in transformed regressor $TZ(t)$ is hard to discuss since it depends on several aspects, where the biggest one is the choice of wavelet filters family.

7.1.1 Case Study

After testing of the algorithm on the simple SISO system in Chapter 6, it is not necessary to demonstrate the MISO algorithm for some testing example and we can test it directly on a real life system. Namely, we will model one part of Heating, Ventilation and Air Conditioning (HVAC) system of the building of the CTU in Prague consisting of one zone only. The controlled quantity is the temperature in the zone T_Z $[K]$ and it is affected mainly by three factors:

1. The first factor is indeed the heating accomplished here by water heating circuit. The water flow rate within the heating circuit is constant, hence the temperature of the supply water T_{SW} $[K]$ is then the only control variable. The water heat up the concrete in the ceiling of the zone, by what the air inside the zone is heated up Prívara et al. [2011]. The temperature of the return water T_{RW} $[K]$ is measured, but the temperature of the concrete T_C $[K]$ is not.

2. The second factor is the ambient temperature T_{amb} $[K]$, which affects the zone temperature through both walls and windows.

3. The last factor is a solar radiation $\dot{Q}_{sol}$ $[Js^{-1}]$ affecting the zone temperature mainly through windows. Note that the physical notation $\dot{Q}$ is used since the solar radiation effect is usually understood as an energy gain of the zone.

[1] The collinearity in the regressor matrix can be also caused by the correlation of input and output data, however, such a case can already appear when dealing with SISO SID.

Thermodynamic systems are worth to be described with respect to energies Prívara et al. [2013a]. A simple first principle model of the situation can be written as

$$\begin{aligned} C_C\dot{T}_C &= c_W\dot{m}(T_{SW} - T_C) + c_W\dot{m}(T_{RW} - T_C) + \frac{l_1 A_C(T_Z - T_C)}{R_{CZ}}, \\ C_W\dot{T}_{RW} &= c_W\dot{m}(T_{SW} - T_{RW}) + c_W\dot{m}(T_C - T_{RW}), \\ C_Z\dot{T}_Z &= \frac{l_2 A_C(T_C - T_Z)}{R_{CZ}} + \frac{l_3 A_S(T_{amb} - T_Z)}{R_{AZ}} + l_4\dot{Q}_{sol}. \end{aligned} \tag{7.2}$$

C_C, C_W and C_Z stand for heat capacities of the concrete in ceiling, heating water and zone air, respectively. c_W denotes the specific heat of heating water. $\dot{m}$ denotes constant water flow rate. A_C and A_S stand for heating area of the ceiling and area of the surface between the zone and external environment, respectively. R_{CZ} and R_{AZ} stand for thermal resistances between concrete and air inside the zone and between zone and ambient environment, respectively. $l_\bullet[-] \in (0, 1)$ are loss factors. Note that in practise, all the capacities, resistances, areas and loss factors are not known and are included within the parameters of the model. All these parameters cover several practical phenomena, e.g. the capacity of the zone air covers also capacity of the equipment inside the zone, the ambient temperature affects the zone temperature through both walls and windows etc. The RC (Resistor-Capacitance) network for this situation is depicted in Figure 7.1. The resistances R_1 and R_2 in the figure represent energy losses of the heating water and are inversely proportional to $c_W\dot{m}$, however, they are introduced only for the RC network purposes and will not be needed anymore.

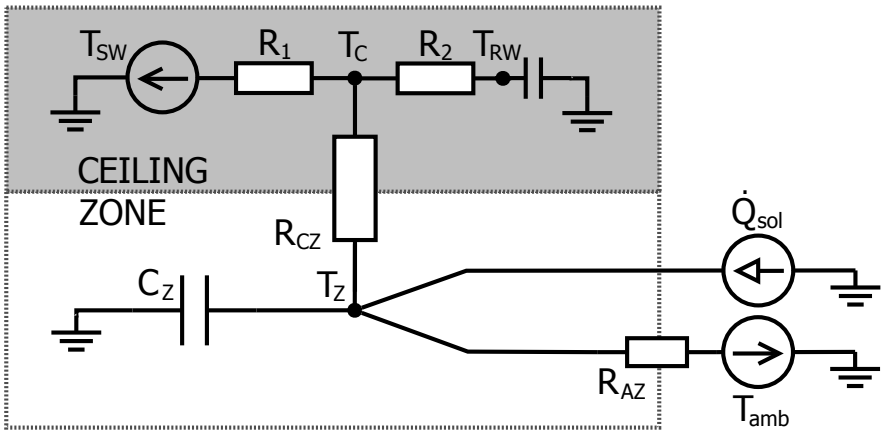


Figure 7.1: A schematic RC network of the heat transfer within the CTU building.

Realize now that from causality point of view, the temperature of return water can not be an input to the system, since it is measured with a delay compared with the effect of the heating water on the zone. From physical point of view, the return water temperature should be an output of the model, however, the system would have been of MIMO structure, which is treated further in the thesis. For the need of MISO structure for this Section, we have to exploit the measured return water temperature

and pre-compute the energy delivered to the concrete (to the point where the T_C is measured) $\dot{Q}_{HEAT}(t) \approx T_{SW}(t-d) - T_{RW}(t+d)$ with d being time delay between T_{SW} and T_C and/or between T_C and T_{RW} and with $\approx$ denoting a proportionality except for an unknown constant. Based on the measurements of supply and return waters, this delay was determined to be approximately 10 minutes, thus $d = 5$ minutes, what is then a reasonable choice for the sampling time T_s to be used for the discretization of the model. Hence $\dot{Q}_{HEAT}(k) \approx T_{SW}(k-1) - T_{RW}(k+1)$ with discrete time k. Moreover, the equation for the $\dot{T}_{RW}$ is now out of the interest. Considering the Euler's approximation of the derivative by a difference, the discrete-time model is now

$$\begin{aligned} C_C \frac{T_C(k+1) - T_C(k)}{T_s} &= l_0 \dot{Q}_{HEAT}(k) + \frac{l_1 A_C (T_Z(k) - T_C(k))}{R_{CZ}}, \\ C_Z \frac{T_Z(k+1) - T_Z(k)}{T_s} &= \frac{l_2 A_C (T_C(k) - T_Z(k))}{R_{CZ}} + \frac{l_3 A_S (T_{amb}(k) - T_Z(k))}{R_{AZ}} + l_4 \dot{Q}_{sol}(k). \end{aligned} \tag{7.3}$$

l_0 covers both a loss factor and an unknown coefficient of proportionality between $\dot{Q}_{HEAT}$ and true energy delivered to the concrete. The model (7.3) is now causal and of the required MISO structure, since the only output of the model is T_Z and the inputs to the model are $\dot{Q}_{HEAT}, T_{amb}$ and $\dot{Q}_{sol}$. The temperature T_C is a state, which however does not play a role within the model's transfer function. The transfer function of the model is (according to (7.3)) of the 2^{nd} order. The parameters $l_\bullet$, $R_\bullet$, $C_\bullet$ and $A_\bullet$ are considered to be constant and together with sampling time T_s form unknown parameters of the model. More details on the modeling of this particular building can be found in Prívara et al. [2011, 2013b]; Široký et al. [2011] (where at some of them, the author of this thesis is a co-author), therefore we will not treat it anymore and let us have a look at the identification using wavelets. For the demonstration, appropriate representative of data measured on the building is depicted in Figure 7.2.

At first, we perform an analysis of data measured on the building. Since the only differences of the temperatures are present in the model description, the temperatures in degrees of Celsius can be used with no loss of generality. In practise, the data usually suffer from several phenomena, what consequently makes their frequency analysis ugly. A quite common issue is the high frequency content of data that is caused by several reasons: i) A common measurement noise is present. ii) A resolution of a sensor causes jumps in the measure values, which consequently add an artificial high frequency content into data. iii) Failures of sensors are similar issue to the letter one. When either the sensor breaks down or the communication fails, it breaks the dynamics stored in the data. And afterwards, when repairing the sensor, another jump in the measured value can arise. We should therefore treat the data carefully, especially from the high frequencies point of view. One of ways to improve the variance properties of the ETFE is to smooth it Ljung [1999]. This procedure results in highlighting of the frequency content of the data across the whole spectra except of the slowest frequencies, since the procedure operates on several (shorter) parts of measured data. Such treatment of the measured data was performed and is depicted in Figure 7.3. Moreover, when performing a frequency analysis of the data, the sampling time $T_s = 5$ minutes emerged

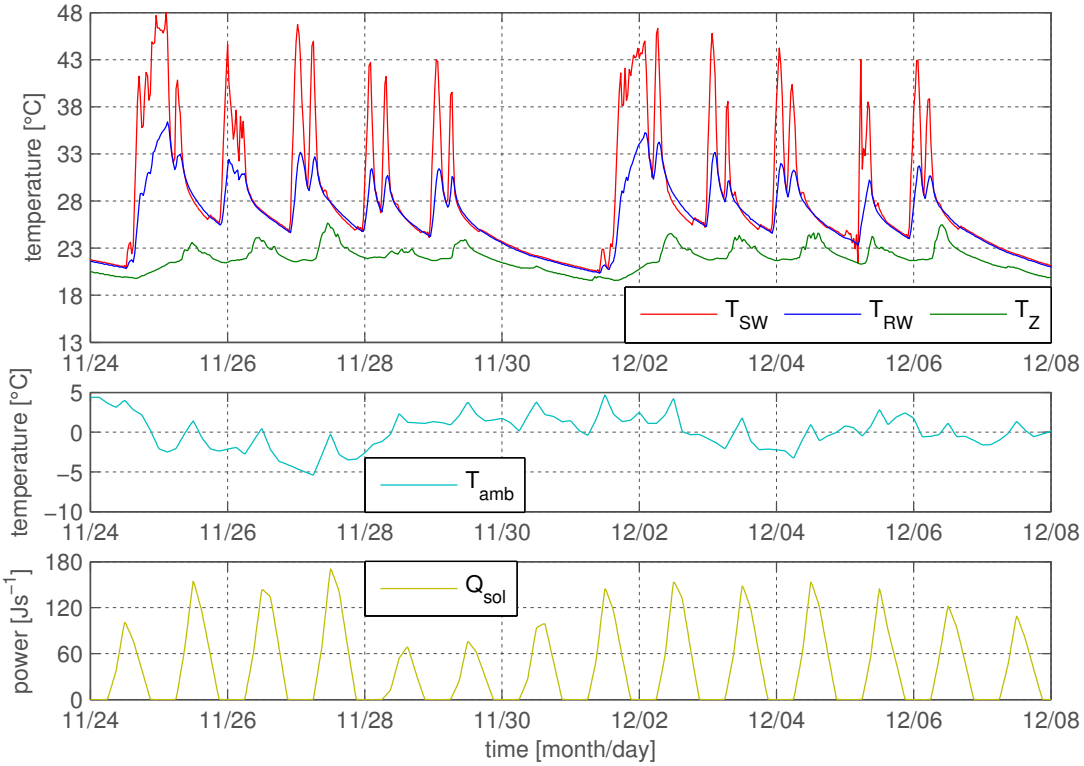


Figure 7.2: Representative data measured on the real building.

to be unnecessarily fast, since the dynamics of the heating up of either the concrete or the air is much slower. Therefore, the $T_s = 5$ minutes was used only for the computation of the energy delivered during one sample $Q_{HEAT} \approx \dot{Q}_{HEAT} \cdot T_s$ and then, the data were re-sampled (or recomputed) onto $T_s = 30$ minutes, what has been proved in literature to be sufficient enough for building modeling.

For the case study, wavelet family db4 was used on the data for identification of length around 800 samples. The maximum analysis level is 6. Figure 7.3 shows the estimates of magnitude-Bode graphs for appropriate inputs, from which the user can presume the importance of individual frequency ranges. Based on the author's experience with building dynamics, the finally used wavelet filters were mainly on the 2^{nd} and on the 3^{rd} level (red and green color; count the levels from the right) together with wavelet filter for steady state. The 1^{st} and the 4^{th} levels (cyan and blue color) should also be considered, since they cover either the fall (poles of the transfer function) or the growth (zeros of the transfer function) in the ETFE. Final vector of weights was set as $V = \begin{bmatrix} 5 & 0 & 0 & 1.5 & 2.3 & 2.3 & 1 \end{bmatrix}$, the orders of the model to be identified was chosen to be $n_a = 2,\ n_b = 3$ and no delay was considered, all in accordance to the transfer function of the model (7.3).

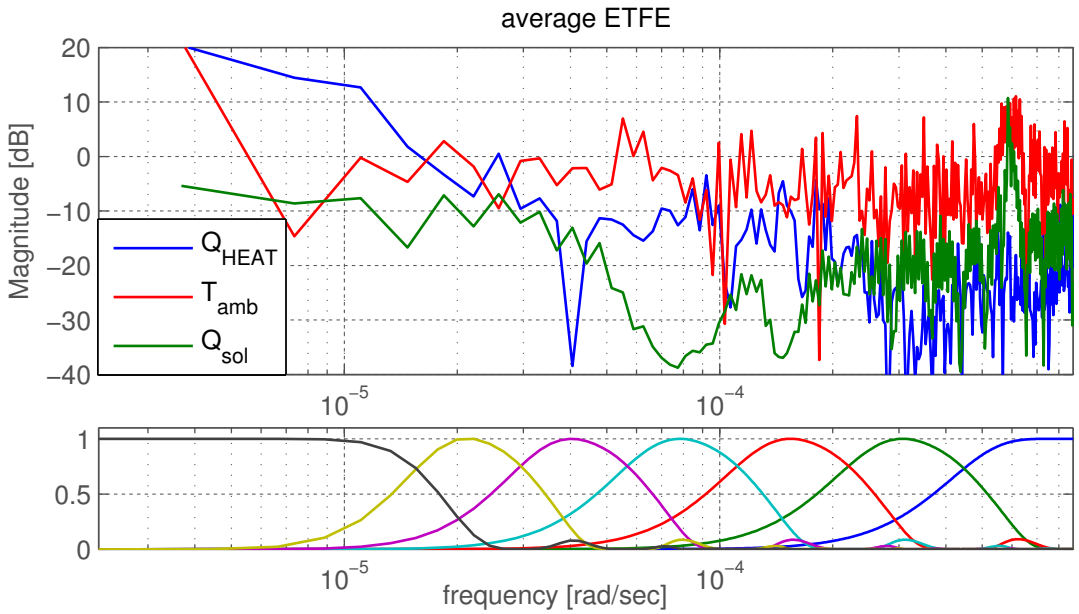


Figure 7.3: Frequency analysis of data measured on the real building.

Figure 7.4 shows the result and possible improvement of the model estimated firstly only by using of the pure ARX model and afterwards by using the wavelets with the above-mentioned settings. Note that the range of the improvement depends (except of the measured data quality) on knowledge of the system. Creating the classical filter is a difficult task in this case since there are several frequencies (or frequency ranges) of interest. On the other hand, the wavelet pre-filtering becomes a very user-friendly method which requires the user to choose the weights for the individual filters only. We can see in that figure that the resulting model possesses acceptable dynamics. Moreover, it is important to realize that, in spite of the model was estimated by PEM, thus with an emphasis to minimize the one step prediction error, Figure 7.4 depicts the simulation of the model on 1 week data,[1] and it still records a good result. On the other hand, comparing the performance of both models on one step ahead prediction would not bring any valuable result, since both models predict the right next output almost exactly (according to the theory and its treatment of the noisy data). Note that to evaluate the simulation performance of the models (e.g. by the well known fit-factor Ljung [1999]) is not so reasonable when simulating the model, since such evaluations usually assess the exact matching of the data, but the emphasis should be laid on the correct dynamical behavior. More detailed comparison of both models is in Table 7.1.

Moreover, we should mention non-linearities, disturbances, noises or even missing measurements in the real life task that also affect the model quality. For example of the buildings modeling, the typical missing variable is an occupancy, which however affects

[1] 1 week simulation is sufficient enough, the MPC control of buildings usually predict the behavior 1 or 2 days ahead Prívara et al. [2013a].

the zone temperature a lot and is also missing in this case study. Finally, Figure 7.4 also depicts histograms of the simulation error. It can be seen, that the WAV model has more Gaussian histogram than the ARX model. It is caused mainly by the satisfactory tracking of the measured output data by WAV model, while the ARX model recorded large simulation error in the second half.

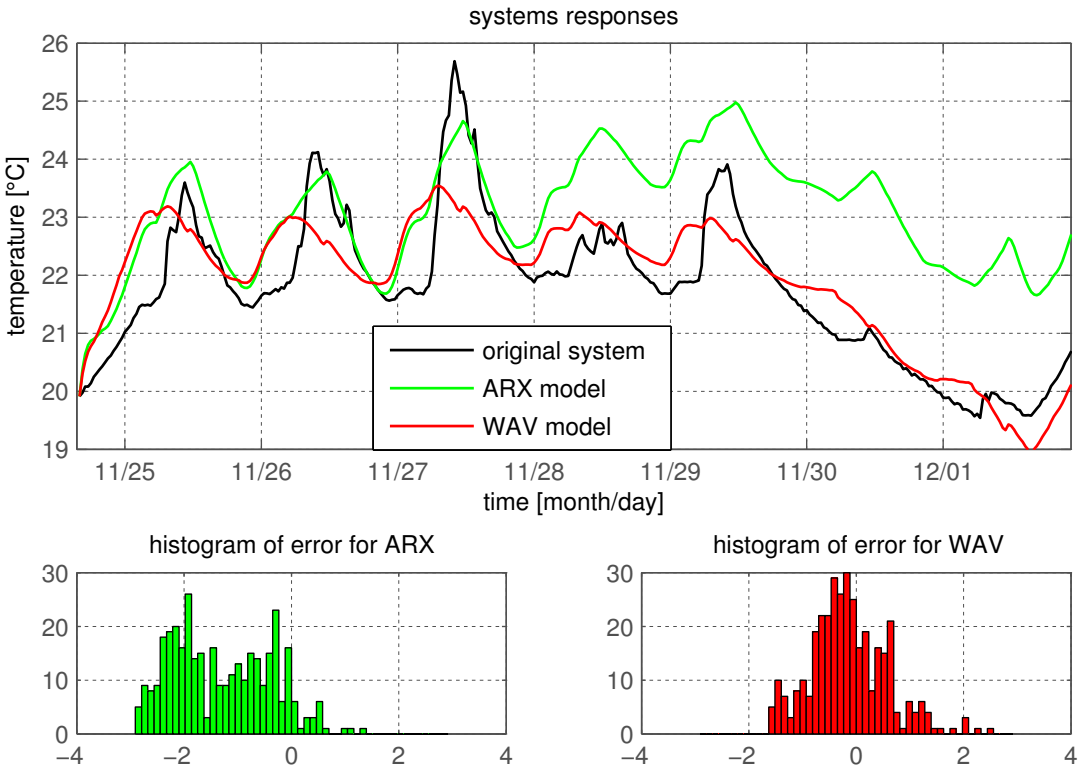


Figure 7.4: Real application of MISO system identification.

7.2 MIMO LTI system identification with wavelets

When facing the problem of modeling and identification of MIMO systems where only input and output data are measured and no or very little information of the system's structure is known, a general model structure has to be utilized. Such a description of the model in the state space form is $x(k+1) = A(\theta)x(k)+B(\theta)u(k)$, $y(k) = C(\theta)x(k)+D(\theta)u(k)$, where dimensions of inputs and outputs are given by measurements, but the order of the system is unknown (the same holds for appropriate dimensions of model matrices). One of the identification methods for such a structure is (already mentioned) 4SID algorithm, which is unsuitable for the DWT incorporation. There is however a possibility to deal with MIMO state space structure in a way enabling the usage of the DWT. This procedure is described in the following Section 7.2.2.

Table 7.1: Comparison of MISO ARX and WAV models.

ARX model	**WAV model**
Fit-factor computed on the simulation	
25.19%	43.21%
Order of the model	
Order $n = 2$, one of them is measured output.	Order $n = 2$, one of them is measured output.
Effect of the inputs	
The ARX model records negative effect of both external inputs what is definitely not correct.	The WAV model records small, thus unreliable, effects of both external inputs what can be a consequence of insufficient inner structure.
Evaluation of the simulation	
The simulation of both models seem to be good enough, especially when considering its length and all the simplifications performed during the modeling. However, despite of the fact that both models possess an acceptable dynamics, effects of their inputs are not correct at all. Therefore, the models are not suitable for any practical utilization, since they (more precisely, their parameters) do not comply basic physical laws.	

In a transfer function description, we can consider a general model described through polynomial fractional matrices $G(q,\theta)$, $H(q,\theta)$. The procedure of identification of such transfer function model structure is described in the following Section 7.2.1, where consequential properties of stochastic part of the model are discussed.

7.2.1 Transfer function description

Assume a multivariable system of the form of (5.1)

$$y(t) = G(q)u(t) + H(q)e(t) \tag{7.4}$$

with

$$G(q) = \begin{bmatrix} G_{1,1}(q) & \cdots & G_{1,m}(q) \\ \vdots & \ddots & \vdots \\ G_{o,1}(q) & \cdots & G_{o,m}(q) \end{bmatrix}, \; H(q) = \begin{bmatrix} H_1(q) & & 0 \\ & \ddots & \\ 0 & & H_o(q) \end{bmatrix}. \tag{7.5}$$

Note that $u(t)$, $y(t)$, $e(t)$ are now vectors. Additionally, since $e(t)$ is considered (as is usual in SID theory) to be white Gaussian and independent sequence, we can "stack" them and write (with some abuse of notation) the stochastic part simply as $H(q) = \begin{bmatrix} H_1(q) & \cdots & H_o(q) \end{bmatrix}^T$. To comply with the theory in Chapter 5, a general model of the

system (7.4) is $y(t) = G(q,\theta)u(t) + H(q,\theta)e(t)$, where $G(q,\theta)$, $H(q,\theta)$ are analogous to those in (7.5).

Let us have a look now at the deterministic part $G(q)$ in more details. From the linear systems theory Antsaklis and Michel [1997], the well-known fact is that any proper rational transfer function $G(q)$ can be written thanks to Polynomial Matrix Fractional Description (PMFD) as

$$G(q) = D_L^{-1}(q)N_L(q) = N_R(q)D_R^{-1}(q), \tag{7.6}$$

where $N_L(q)$, $N_R(q)$, $D_L(q)$, $D_R(q)$ are polynomial matrices of appropriate dimensions. For further purpose of the work, we will use only the notation $G(q) = D_L^{-1}(q)N_L(q) = D^{-1}(q)N(q)$. The matrix $D(q)$ can be computed as a diagonal matrix with elements as the least common denominators of all entries in the appropriate rows of $G(q)$

$$D(q) = \begin{bmatrix} d_1(q) & & 0 \\ & \ddots & \\ 0 & & d_o(q) \end{bmatrix}, \tag{7.7}$$

however, in such configuration, the diagonal elements $d_i(q), i = 1, \ldots, o$ can be different, so the MIMO system can be split into o MISO systems and each identified separately:

$$\begin{aligned} d_1(q)y_1(t) &= N_1(q)u(t), \\ \vdots \qquad & \qquad \vdots \\ d_o(q)y_o(t) &= N_o(q)u(t), \end{aligned} \tag{7.8}$$

where N_i stands for the i^{th} row of matrix $N(q)$. This structure has many advantages Zhu [2001]: i) It has simple interpretation. ii) Splitting MIMO model into MISO models reduces a complexity of the SID problem. iii) The SID methods developed for SISO systems (see Chapter 5) and extended to MISO systems (see Section 7.1) can be further used for MIMO systems as well.

Having now $G(q) = D^{-1}(q)N(q)$ in the aforementioned form, we obtain the model structure

$$y(t) = D^{-1}(q,\theta)N(q,\theta)u(t) + H(q,\theta)e(t). \tag{7.9}$$

The choice for the structure of $H(q,\theta)$ should be done to classify the SID problem. It is easy to see that when $H(q,\theta) = I$, the model (7.9) becomes multivariable OE model structure consisting of o independent MISO OE models. Another choice is $H(q,\theta) = D^{-1}(q,\theta)$ which yields the MISO ARX model structure

$$D(q,\theta)y(t) = N(q,\theta)u(t) + e(t). \tag{7.10}$$

This choice gives us linear in parameters predictor $\hat{y}(t) = G(q,\theta)u(t) = (I - D(q,\theta))y(t) + N(q,\theta)u(t)$. Each partial MISO system is then identified as in Section 7.1.

Another option is to choose the matrix $D(q)$ as $D(q) = d(q)I$ with polynomial $d(q)$ being a least common denominator of all entries in $G(q)$, what consequently means that all partial transfer functions in

$$G(q) = D^{-1}(q)N(q) = \frac{N'(q)}{d(q)} \tag{7.11}$$

have the same denominator, $N'(q)$ stands for $N(q)$ with some polynomials extended in order to the equality (7.11) holds. The model structure (7.10) now has an expression

$$\begin{aligned} d(q,\theta)y_1(t) &= N_1(q,\theta)u(t) + e_1(t), \\ &\vdots \qquad\qquad \vdots \\ d(q,\theta)y_o(t) &= N_o(q,\theta)u(t) + e_o(t). \end{aligned}$$

As the denominator $d(q,\theta)$ is common to the whole model, the user can identify the whole system at once. Moreover, it enables the user to select the order of the denominator according to considered important dynamics of the system.

The formulation of the identification problem follows. Predictor of the i^{th} MISO submodel can be written according to Section 7.1 as $\hat{Y}_i(t,\theta) = Z_i(t)\theta_i$, where θ_i contains unknown parameters of $d(q,\theta)$ and $N_i(q,\theta)$. Since the part of θ_i containing parameters of $d(q,\theta)$ is common to all θ_i, $i = 1,\ldots,o$, let us divide $Z_i(t)$ and θ_i as $Z_i(t) = [Z_{i,d}(t), Z_{i,N}(t)]$ and $\theta_i = \left[\theta_d^T, \theta_{i,N}^T\right]^T$, where θ_d^T and $\theta_{i,N}^T$ contains only the parameters of $d(q,\theta)$ and $N_i(q,\theta)$, respectively, $Z_i(t)$ is divided in the same manner. Now the predictor for the whole MIMO model can be written in the form of $\hat{Y}(t,\theta) = Z(t)\theta$ as

$$\underbrace{\begin{bmatrix} \hat{Y}_1(t,\theta) \\ \hat{Y}_2(t,\theta) \\ \vdots \\ \hat{Y}_o(t,\theta) \end{bmatrix}}_{\hat{Y}(t,\theta)} = \underbrace{\left[\begin{array}{c|c|c|c|c} Z_{1,d}(t) & Z_{1,N}(t) & 0 & \cdots & 0 \\ \hline Z_{2,d}(t) & 0 & Z_{2,N}(t) & & \vdots \\ \hline \vdots & \vdots & & \ddots & 0 \\ \hline Z_{o,d}(t) & 0 & \cdots & 0 & Z_{o,N}(t) \end{array}\right]}_{Z(t)} \underbrace{\left[\begin{array}{c} \theta_d \\ \hline \theta_{1,N} \\ \hline \theta_{2,N} \\ \hline \vdots \\ \hline \theta_{o,N} \end{array}\right]}_{\theta}.$$

Obviously, the problem can now be solved via OLS as in (6.2) with $Y(t)$ being compound in the same manner as $\hat{Y}(t,\theta)$.

The main pros and cons of this approach with $D(q,\theta) = d(q,\theta)I$ compared to the approach with general diagonal matrix $D(q,\theta)$ are:

+ The whole model is identified at once.

+ The number of unknown parameters is significantly reduced, since the only one denominator is used for the whole model.

– $Y(t)$, $Z(t)$ matrices used within the identification problem are much larger, thus the computational demands increased. However, the matrix $Z(t)$ is very sparse and has specific structure, what can simplify the computation.

The presented method extends the MISO SID from the Section 7.1, therefore, the properties of the method remain the same as before as well as issues with correlation of data or collinearity in regressor. Finally, the way of utilizing wavelets within this method still remains as before with one important, but quite clear change: each partial MISO model can be obtained via a different set of wavelet filters, hence with emphasis placed on different dynamics.

Concerning the utilization of different model structure of MIMO system (or MISO systems) like aforementioned ARMAX or OE model structures, the difficulty of exploiting a recursive identification together with wavelets has already been discussed within Section 6.5 and the same holds for a multivariable case as well.

7.2.2 State space description

Consider now the true system is MIMO LTI system in a state space description with all outputs being also the state variables and that all of them are measured. The discrete formulation of such a system is

$$x(k+1) = Ax(k) + Bu(k) + e(k)$$

with k being the discrete time, $x \in \mathbb{R}^n,\ u \in \mathbb{R}^m,\ e \in \mathbb{R}^n$ are system state (or output), input and measurement noise vectors, respectively. Matrices A and B are system matrices of appropriate dimensions. The state space model structure can be chose as

$$x(k+1) = A(\theta)x(k) + B(\theta)u(k) \tag{7.12}$$

with $A(\theta),\ B(\theta)$ being model matrices. For the time line, it can be written

$$X_2^N = A(\theta)X_1^{N-1} + B(\theta)U_1^{N-1}, \tag{7.13}$$

where N is the number of samples and X_k^{k+N-1}, U_k^{k+N-1} are the matrices of state, input and noise values defined as follows

$$\begin{aligned} X_k^{k+N-2} &= \begin{bmatrix} x_k & x_{k+1} & \dots & x_{k+N-2} \end{bmatrix}, \\ U_k^{k+N-2} &= \begin{bmatrix} u_k & u_{k+1} & \dots & u_{k+N-2} \end{bmatrix}. \end{aligned} \tag{7.14}$$

The equation (7.13) can be rewritten as

$$X_2^N = \begin{bmatrix} A(\theta) & B(\theta) \end{bmatrix} \begin{bmatrix} X_1^{N-1} \\ U_1^{N-1} \end{bmatrix}. \tag{7.15}$$

For the purpose of possibility to apply the DWT within this SID method, the (7.15) has to be transposed. However, due to a simple notation, we set up the notation $X_2 = X_2^N,\ X_1 = X_1^{N-1},\ U_1 = U_1^{N-1}$ and then obtain

$$X_2^T = \begin{bmatrix} X_1^T & U_1^T \end{bmatrix} \begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}.$$

This equation can be further rewritten to the standard OLS estimation problem with the aid of vectorization (vec(•)) and Kronecker product (• ⊗ •)

$$\operatorname{vec}\left(X_2^T\right) = \left(I_n \otimes \begin{bmatrix} X_1^T & U_1^T \end{bmatrix}\right) \operatorname{vec}\left(\begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}\right), \tag{7.16}$$

where I_n is $n \times n$ identity matrix and n represents the system's order. The equation (7.16) is now in the form of $Y = Z\theta$ with

$$Y = \operatorname{vec}\left(X_2^T\right) \ , \ Z = I_n \otimes \begin{bmatrix} X_1^T & U_1^T \end{bmatrix} \ , \ \theta = \operatorname{vec}\left(\begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}\right).$$

Let us have a look at the predictor for the model (7.12). The transfer functions of this model are $G(q,\theta) = \left(q^{-1}I - A(\theta)\right)^{-1} B(\theta)$, $H(q,\theta) = I$ and the predictor (in order to (5.5)) is

$$\begin{aligned} \hat{Y}_2(\theta) &= \left[I - H^{-1}(q,\theta)\right] Y_1 + H^{-1}(q,\theta) G(q,\theta) U_1 \\ &= \left[I - I^{-1}\right] Y_1 + I\left(q^{-1}I - A(\theta)\right)^{-1} B(\theta) U_1 \\ &= \left(q^{-1}I - A(\theta)\right)^{-1} B(\theta) U_1, \end{aligned}$$

with Y_1, Y_2 being in accordance to (7.14). Therefore, since $\left(q^{-1}I - A(\theta)\right)^{-1} B(\theta)$ is the transfer function and since the outputs are also the state variables, we can write the predictor in the form of

$$\hat{X}_2(\theta) = A(\theta) X_1 + B(\theta) U_1$$

and due to (7.13), $\hat{X}_2(\theta) = X_2(\theta)$. The consequence is that the equation (7.16) is predictor of the model (7.12) and $\hat{Y} = \operatorname{vec}\left(\hat{X}_2^T\right) = \operatorname{vec}\left(X_2^T\right) = Y$.

The problem of finding the parameters θ in the prediction error minimization point of view is now well known and corresponds to the (6.2) and (6.3). So in our case it can be written as

$$\begin{aligned} \hat{\theta}_N &= \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \left[X_2 - \hat{X}_2(\theta)\right]^2 \\ &= \arg\min_{\theta} \left\| \operatorname{vec}\left(X_2^T\right) - \left(I_n \otimes \begin{bmatrix} X_1^T & U_1^T \end{bmatrix}\right) \operatorname{vec}\left(\begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}\right) \right\|_2^2 . \end{aligned} \tag{7.17}$$

An alternative expression is

$$\hat{A}_N(\theta), \hat{B}_N(\theta) = \arg\min_{A(\theta), B(\theta)} \left\| \operatorname{vec}\left(X_2^T\right) - \left(I_n \otimes \begin{bmatrix} X_1^T & U_1^T \end{bmatrix}\right) \operatorname{vec}\left(\begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}\right) \right\|_2^2 \tag{7.18}$$

$$\begin{aligned} \text{s.t.:} \quad & A(\theta) \in \mathcal{A}_{\mathcal{M}}, \\ & B(\theta) \in \mathcal{B}_{\mathcal{M}}, \end{aligned}$$

where $\mathcal{M}$ stands for the particular model structure including both the structure of matrices $A(\theta)$, $B(\theta)$ and constraints of parameters θ and where $\mathcal{A}_\mathcal{M}, \mathcal{B}_\mathcal{M}$ are sets of all admissible matrices $A(\theta)$, $B(\theta)$. The difference between these two expressions is that the prediction error in (7.18) is linear in parameters $A(\theta)$, $B(\theta)$ or their elements, respectively, but the prediction error in (7.17) need not be linear in θ. The solvability of (7.18) depends on the choice of what the parameters θ are. Let us explain it quickly on an example. Let $A = \begin{bmatrix} ab & c \\ 0 & b^2 \end{bmatrix}$, where a, b, c are unknown. When $\theta = [ab, c, b^2]$, the problem is the same as the (7.18) and is uniquely solvable. On the other hand, when $\theta = [a, b, c]$, we have to introduce an auxiliary $\theta' = f(\theta) = [ab, c, b^2]$ to solve the problem as before, but existence of an analytical solution to $\theta = f^{-1}(\theta')$ is not guaranteed, not even its uniqueness.

Realize now that when writing the model structure based on physics, only some of elements of matrices $A(\theta), B(\theta)$ are to be found while the others can be zero. This is the point where the strength of the vectorization comes up. Namely, the set of equations (7.16) can be divided on the equations with zero right side and the equations with right side dependent on θ. The zero-right-side equations can then be omitted from the problem (7.17), since they are independent of parameters θ and they do not affect the solution $\hat{\theta}_N$. Moreover, the selected (modelled based on physics) inner structure of matrices $A(\theta), B(\theta)$ can be verified by evaluating the errors of the zero-right-side equations and, if necessary, an analysis of missing terms (missing physical phenomena) can be performed.

Note that there is another option that some parameter is known. Then the only difference from the principle described above is that the set of equations with right side independent on θ extends. On the other hand, the user need not trust this value, so leave it among unknown parameters as one of them and either initiate it on the assumed value or limit it by some reliable range around the assumed value.

It can be seen on the $\begin{bmatrix} X_1^T & U_1^T \end{bmatrix}$ matrix that the transposition preserved the time-structure of the data. Moreover this time-structure is not spoiled by the Kronecker product $\left(I_n \otimes \begin{bmatrix} X_1^T & U_1^T \end{bmatrix}\right)$, hence such the SID problem formulation is suitable for the incorporation of the DWT. The wavelet filtering can be incorporated directly into the minimization problem (7.17) in many ways, for example as

$$\hat{\theta}_N = \arg\min_\theta \left\| \mathrm{vec}\left(WTX_2^T\right) - \left(I_n \otimes WT \begin{bmatrix} X_1^T & U_1^T \end{bmatrix}\right) \mathrm{vec}\left(\begin{bmatrix} A^T(\theta) \\ B^T(\theta) \end{bmatrix}\right) \right\|_2^2. \tag{7.19}$$

where T, W are wavelet and weighting matrix, respectively. Note that the principle of interconnection of the DWT and the SID again remains the same is before.

7.2.3 Case Study

In here, we will consider another case study to show a functionality of the proposed algorithm. Firstly, we assume a system

$$G(q) = \begin{bmatrix} \frac{2s/30+1}{s^2+2s+1} & \frac{10s/3+100}{s^2+20s+100} & \frac{50s+100}{s^2+20s+100} \\ \frac{0{,}125s+10}{s^2+11s+10} & \frac{s/30+1}{s^2+11s+1} & \frac{2s+10}{s^2+11s+10} \end{bmatrix},$$

and discretize it by $T_s = 0.01$ s. This system has two main locations of poles (in -1 and in -10), fast and slow, and we will show just the result of applying the algorithm for MIMO SID, since the idea of applying the wavelets has been discussed in more details within case studies in Chapter 6 and Section 7.1. The system was excited at all frequencies by Pseudo-Random Binary Sequence (PRBS), data used for the identification were of the length of 2000 samples and basic wavelet function db5 was used. Since there is higher number of equation containing the high-frequencies information, we will focus on a fast submodel identification only (so the frequencies corresponding to $\omega = 10$ and higher). The original system is of 2^{nd} order and the order of the model to be identified was chosen as $n_a = 1, n_b = 2$ and with no delay. The resulting model covers the high frequencies only, what is demonstrated by comparison of Bode plots of original system and the resulting model in Figure 7.5.

Another illustration was performed on a real life case. In the case study for MISO system (Section 7.1.1) we promised that the full causal structure of the building of the CTU in Prague will be treated as a MIMO system, so it is here. Based on the equations (7.2) and proper causal sampling, we obtain the 3^{rd} order model with T_{SW}, T_{amb} and $\dot{Q}_{sol}$ as inputs, T_C as unmeasured state and T_{RW} and T_Z as measured states, thus outputs. Note that two cases will be showed here:

1. Since one of states of the model is not measured, the identification method for the MIMO state-space model description (Section 7.2.2) can not be used and we use the transfer function approach (Section 7.2.1). This case will be treated in a similar manner as was in the MISO case in Section 7.1.1, thus we will compare the pure ARX model against the WAV model exploiting the wavelets.

2. Next, we add the comparison against the model, which is widely used in practise of the predictive control of the building heating system Prívara et al. [2013a]; Váňa et al. [2014]. Such a model is identified in the form of state-space description via the method described in Section 7.2.2. The state-space description of the model is enabled by the omitting of the virtual concrete core temperature T_C. In this case, the model can obviously be of the lower order. This model is further marked as GB model.

The same identification data as well as the same type of wavelet filters as for the MISO case were used. It means the data length of around 800 samples were used, what with wavelet function db4 results in that the maximum wavelet analysis level is 6.

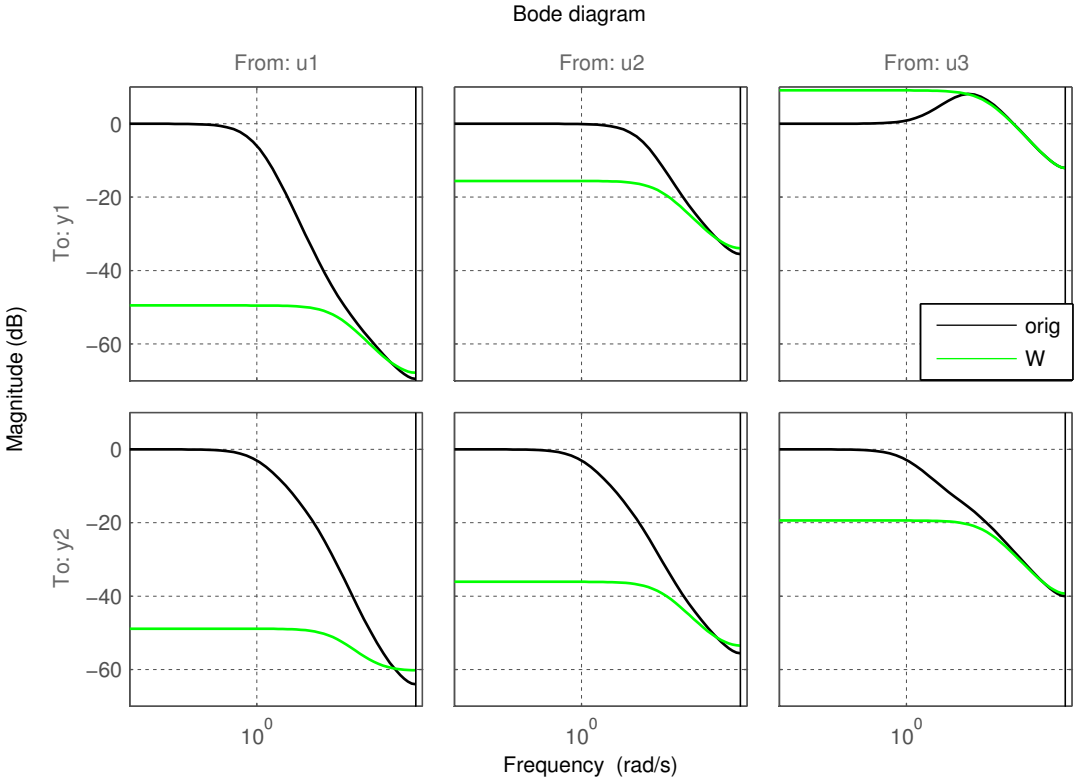


Figure 7.5: Comparing of Bode plots of the resulting MIMO fast submodel and the system.

The order of the model must comply the previously mentioned model structure, thus $n_a = 3, n_b = 4$ and again, no delay was considered. The only difference in the procedure is now in the set of inputs and outputs of the model as their mutual relationships are different than for the MISO case. Therefore, there will be a different vector of weights. Based on the author's experience with building modeling, the weights were set as a combination of two approaches to weighting:

1. The first approach is weighting the number of equations within a particular frequency range (see Section 6.3). Since each new analysis level halves (approximately) the number of equations from the analysis on one level lower, the vector of weights were set as $V_1 = \begin{bmatrix} 64 & 32 & 16 & 8 & 4 & 2 & 1 \end{bmatrix}$. It means that the cumulative errors at all distinct frequency ranges are of the same interest.

2. The second approach is classical weighting of frequency ranges of interest. The vector of weights was set as $V_2 = \begin{bmatrix} 12 & 1 & 3 & 2 & 3 & 1 & 3 \end{bmatrix}$.

The final vector of weights is then element-wise multiplication of V_1 and V_2.

The results are introduced in the same manner as in Section 7.1.1. However to be more interesting, the simulation was performed for much longer time interval. Figure 7.6

and Figure 7.7 show the 6-weeks simulation of all resulting models in comparison both to the measured data and to each other, each figure for one particular output. Based on the length of the simulation, there can be no doubts that all models possess a proper dynamics. Realize that since the objective of the application is to control the temperature T_Z inside the zone, an emphasis during identification should be placed right on this output. Moreover, as is visible from the Figure 7.6, the models perform very well from the T_{RW} point of view. Therefore, we can allow some small lost of performance on this output at the expense of improvement of performance on the zone temperature T_Z output. Such a user's endeavour can be seen on histograms of errors of ARX and WAV models, where for T_{RW}, the ARX model recorded better result than the WAV model while for T_Z, the WAV model recorded significantly better result than the ARX model. There is another important result visible from histograms so that there are no systematic errors (as a shift) in the simulation, what also verifies the qualities of the models.

Concerning the GB model, it recorded results which are comparable to those of the ARX and the WAV models, even despite of the fact that it is of the lower order. In the predictive control, the lower model order means also the lower computational demands. On the other hand, the difference of the orders of the models is not that high to cause any computational difficulties when using for the predictive control of buildings, since the heating has slow dynamics and there is a lot of time to compute the control action. Yet another fact is that the state-space identification approach (used for obtaining of the GB model) is sufficient enough to describe the system's behavior, since it is widely used for the control in practise Prívara et al. [2013b]; Váňa et al. [2014]. More detailed comparison of both models is in Table 7.2.

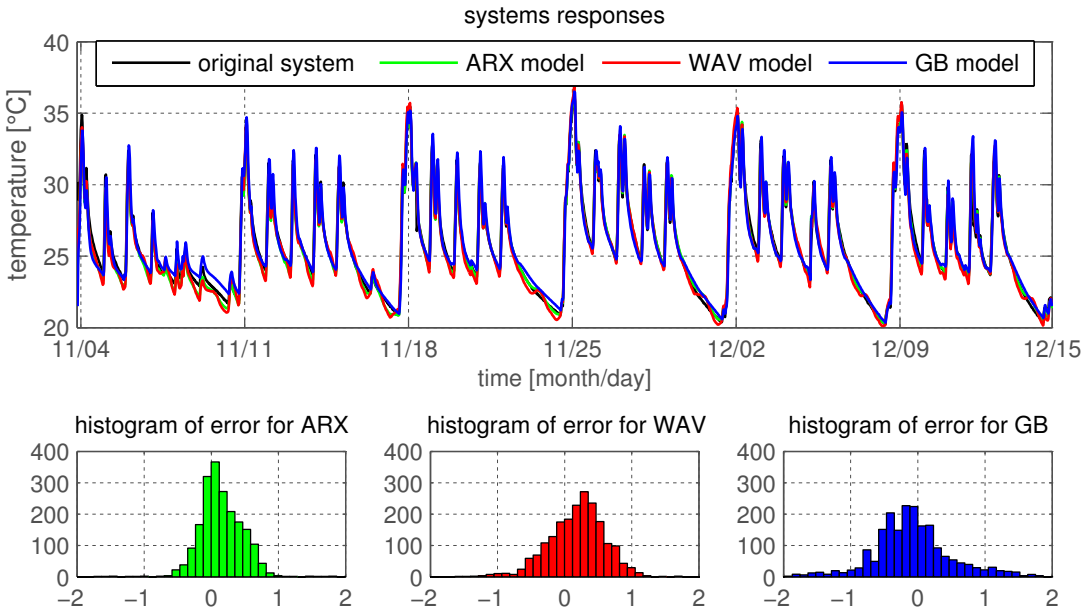


Figure 7.6: MIMO real life case study - a comparison of models simulations against the measured data - **the return water temperature** behavior.

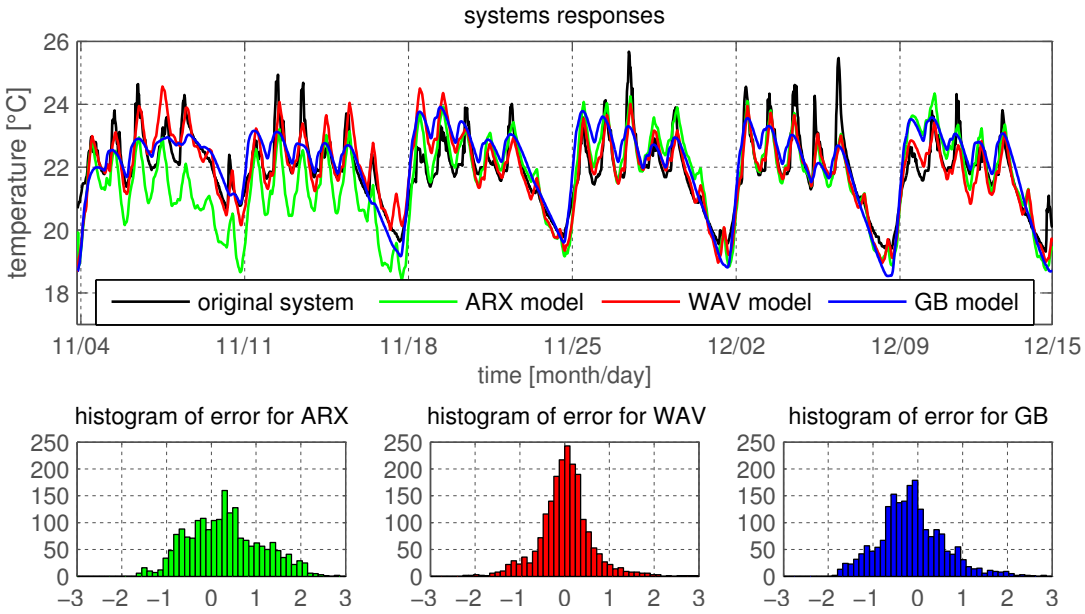


Figure 7.7: MIMO real life case study - a comparison of histograms of simulated output errors - **the zone temperature** behavior.

Table 7.2: Comparison of MIMO ARX, WAV and GB models.

ARX model	**WAV model**	**GB model**
Fit-factor computed on the simulation of T_{RW}		
90.19%	86.72%	81.51%
Fit-factor computed on the simulation of T_Z		
24.52%	49.70%	35.23%
Order of the model		
Order $n_{min} = 6$, none of them is measured. Moreover, the model contains unjustified oscillatory dynamics.	Order $n_{min} = 6$, two of the are measured outputs. There are 3 main eigenvalues of the state transition matrix repeating for each output.	Order $n_{min} = 2$, both of them are measured outputs. The eigenvalues are similar to those of the WAV model, however the fastest dynamics for the heating water itself is missing.
Effect of the inputs		
The ARX model records negative effect of the solar input what is definitely not correct.	The WAV model records (out of the considered models) the most reliable effects of all inputs.	The GB model records reliable effects of inputs on T_Z, but the effects on the T_{RW} seem to be overestimated.
Evaluation of the simulation		
The simulation of all models seem to be good enough, especially when considering its length.		
Despite of the fact that the ARX model possesses an acceptable dynamics, effects of its input is not correct, therefore, the ARX model is not suitable for any practical utilization, since it does not comply basic physical laws.	The WAV model records very good long-time simulation result, even better than the GB model, what is caused by both the higher order and stronger emphasis on important frequencies during the identification. Moreover, not measuring of some states is not such a big deal since a state observer is a common solution to this issue in practise.	The GB model records very good long-time simulation result. Considering the simplicity of the model together with advantage of the measured states, it definitely belongs among the models suitable for further practical MPC control.

8 Wavelets as modulating functions

According to practical data measurement, parameters estimation from the discrete-time domain is the dominating approach in system identification. On the other hand, models based on physical laws are usually described in the continuous-time domain, what is a motivation for using identification methods for continuous-time models.

There are several approaches to continuous-time system identification Garnier et al. [2003], many of them are well-established and widely known. They divide into exact and direct methods. The exact approach exploits the discretization of original system to estimate its parameters, namely, the discrete-time model is described in terms of continuous-time parameters, which are then computed from parameters estimated for the discrete-time model. In direct approach, continuous-time signals are transformed in order to obtain a set of algebraic equations, whose solving results in required parameters.

This chapter describes one of the basic identification methods developed mainly for continuous-time model estimation together with possible incorporation of wavelet functions. The state of the art for this chapter has already been stated in Chapter 3.

8.1 Modulating function method

Consider the continuous-time LTI system

$$y^{(n_a)}(t)+a_{n_a-1}y^{(n_a-1)}(t)+\cdots+a_0y(t) = b_{n_b}u^{(n_b)}(t)+b_{n_b-1}u^{(n_b-1)}(t)+\cdots+b_0u(t)+e(t),$$

where $n_b \leq n_a$, where $e(t)$ is a noise and where $a_0, a_1, \ldots, a_{n_a}, b_0, b_1, \ldots, b_{n_b}$ are parameters to be estimated from a set of input and output signals $y(t), u(t)$. $\bullet^{(i)}$ stands for the i^{th} derivative of the function with respect to time. For a simple notation, we can write

$$\sum_{i=0}^{n_a} a_i y^{(i)}(t) = \sum_{j=0}^{n_b} b_j u^{(j)}(t) + e(t) \ , \ a_{n_a} = 1. \tag{8.1}$$

The fundamental complication of this formulation is within the time derivatives. There are three methods to cope with this issue Garnier et al. [2003] and the approach using

modulating functions is one of them.

The basic principle of applying modulating function for SID was firstly suggested by Shinbrot [1954], however for more details the reader can look for some recent book, e.g. Rao and Unbehauen [2006]. The idea of modulating functions is quite simple as they are only weighting functions to the signals. The only (and the main) difference is that they become zero-valued after a finite time. The reason for this feature is best explained by applying the procedure.

Modulating functions $\phi(t)$ of general order n are smooth, n times differentiable functions with finite support $[0, T]$. Mathematically, they satisfy 3 specific properties:

1.
$$\phi(t) = \begin{cases} \phi(t) & \text{for } 0 \leq t \leq T, \\ 0 & \text{elsewhere.} \end{cases} \tag{8.2}$$

2.
$$\frac{d^i \phi}{dt^i}(t) = \phi^{(i)}(t) \text{ exists for } i = 0, \ldots, n. \tag{8.3}$$

3.
$$\phi^{(i)}(0) = 0 \ , \ \phi^{(i)}(T) = 0 \text{ for } i = 0, \ldots, n. \tag{8.4}$$

The main purpose of defining the modulating functions as above is their consequent property:

$$\int_0^T f^{(i)}(t)\phi(t)dt = (-1)^i \int_0^T f(t)\phi^{(i)}(t)dt, \tag{8.5}$$

where $f(t)$ is a continuous-time n times differentiable function. Note that this important property can be derived by repeated integrating by parts together with using all three properties of modulating functions.

Let us apply it onto the identification of the system (8.1). Modulating (multiplying) the system by the function $\phi(t)$ and integrating over the time interval $[0, T]$ yields

$$\sum_{i=0}^{n_a} a_i \int_0^T y^{(i)}(t)\phi(t)dt = \sum_{j=0}^{n_b} b_j \int_0^T u^{(j)}(t)\phi(t)dt + \int_0^T e(t)\phi(t)dt \ , \ a_{n_a} = 1.$$

However, because of the property (8.5), the expression above can be rewritten as

$$\sum_{i=0}^{n_a} (-1)^i a_i \int_0^T y(t)\phi^{(i)}(t)dt = \sum_{j=0}^{n_b} (-1)^i b_j \int_0^T u(t)\phi^{(j)}(t)dt + \int_0^T e(t)\phi(t)dt \ , \ a_{n_a} = 1.$$

Now, it is possible to evaluate the integrals numerically, what leads to the sole algebraic equation. Hence the modulating functions enable us to transform the continuous-time problem into set of algebraic equations and, moreover, they avoid the effect of initial conditions when estimating unknown system parameters. The number of unknown parameters is $n_a + n_b + 1$, so at least the same number of linearly independent algebraic equations is needed to estimate the parameters. However, having exactly $n_a + n_b + 1$

equations need not be satisfying, since the power of statistical data-driven SID is in the much larger amount of equations than of unknown parameters and since it is hard to obtain really independent equations in practise. There are two usual ways of generating the equations:

1. When $\phi(t)$ is the modulating function, the shifted function $\phi(t - \Delta T)$ is also the modulating function. The only difference is that its support is $[\Delta T, T + \Delta T]$, but the basic principle of modulating function method is preserved.

2. The second and an obvious alternative is to use a different modulating function.

Denote the particular modulating function as $\phi_k(t)$ and denote

$$\gamma_k^i(y) = (-1)^i \int_0^T y(t)\phi_k^{(i)}(t)dt, \tag{8.6}$$

$$\epsilon_k = \int_0^T e(t)\phi_k(t)dt. \tag{8.7}$$

Then the algebraic equation for some $\phi_k(t)$ can be rewritten as

$$\gamma_k^{n_a}(y) = -\sum_{i=0}^{n_a-1} a_i\gamma_k^i(y) + \sum_{j=0}^{n_b} b_j\gamma_k^j(u) + \epsilon = \gamma\theta + \epsilon,$$

where

$$\gamma = \left[-\gamma_k^{n_a-1}(y), \ldots, -\gamma_k^0(y), \gamma_k^{n_b}(u), \ldots, \gamma_k^0(u)\right], \tag{8.8}$$

$$\theta = \left[a_{n_a-1}, \ldots, a_0, b_{n_b}, \ldots, b_0\right]^T. \tag{8.9}$$

Matrix extension of this equation for all employed modulating functions $\phi_k(t)$ is solvable via classical OLS method.

For more details about not only the modulation functions method itself, but about the application of this method on several issues concerning the system identification as well, please, see Preisig and Rippin [1993a].

8.1.1 Choice of modulating function

As the modulating functions are the only variables in this method, they should be selected responsibly to correspond to the aims of the identification procedure = the purpose of the model. The selection should be done in order to the following properties of the modulating functions:

1. **Type:** Several types of modulating functions can be found across the literature. The widely referred types are Shinbrot's functions, Cahen's and Loeb's functions and Maletinsky's spline functions, however, many other successful attempts to create the modulating function were done. Except the properties discussed further, they usually differ in number of parameters.

2. **Order:** The main rule which the order of the modulating function should satisfy is that it is larger than or equal to the order of the identified model, thus to be at least n times differentiable.

3. **Time properties:**

   i) **Length:** It has been already shown that the finite (compact) support is essential property of modulating function. The length T of the modulating function obviously determines the lowest distinguishable frequency in its spectra. Also the maximum length is limited by the length of the measured data. Next realize that the data are always corrupted by a noise, which is integrated over T as well. So the choice of T should be some trade-off between stochastic part elimination and extraction of useful information from the data.

   ii) **Shift:** The shift ΔT plays an important role when generating algebraic equations needed for problem solving. Since the particular equations are beneficial to be independent, the suitable choice for time shift has to be done. The choice itself is not easy, since it can depend on the choice of modulating function type or even on the excitation of the data (or the system). Moreover, several various shifts $\Delta T_k,\ k = 1, 2, \ldots$ can be used for different modulating functions. Choice for proper ΔT is sometimes mentioned in connection with the frequency information obtained from data, realize however that the shift has no impact on the modulating function spectra.

4. **Frequency properties:** Since the modulating functions are usually symmetrical around some time point t_s (mostly due to their simple construction and consequently for the fast computation), the equality

$$\begin{aligned}\int_0^T f^{(i)}(t)\phi(t_s - t)dt &= \int_0^T f^{(i)}(t)\phi(t_s + t)dt \\ &= (-1)^i \int_0^T f(t)\phi^{(i)}(t_s + t)dt \\ &= (-1)^i \int_0^T f(t)\phi^{(i)}(t_s - t)dt\end{aligned}$$

   holds. It transfers the multiplying with shifted modulating function (by t_s) into the convolution in time point t_s and as is generally known, the convolution in time is multiplying in frequency, thus filtering. The conclusion is that because of the modulating functions are symmetrical, the modulating can be thought of as a simple filtering. Therefore, the user should now take care about the frequency properties of modulating functions as well. Some of them are already mentioned above within the time properties, so let them repeat from the frequency point of view. The maximum distinguishable frequency is given by the length T (or by the length of data) together with the Shannon-Kotelnik theorem. Also various types of modulating function can differ in their spectra, thus each type can be

utilized to focus on an appropriate (application depended) frequency information in data.

Next, the stochastic part is filtered as well, so when the noise is (rightfully) considered to be in higher frequencies, we can practically get rid of the noise by proper choice of modulating functions ans reduce the error in parameters estimation significantly. However, the general conclusion about the stochastic properties of the estimate is hard to make since it finally depends on many user's choices, which each treat the noise diversely.

8.2 Wavelets as modulating functions

When reading about the modulating functions, several similarities to wavelet transform arise, e.g. requirement on finite support or specific time-frequency localisation. This section introduces utilizing wavelet functions as modulating functions within the system identification procedure. First, the relationship between wavelets and modulating functions is determined and then, a discussion on both choice of wavelets and other aspects related to wavelet analysis is provided here.

8.2.1 Wavelets and modulating functions relationship

As the modulating function method is method for continuous-time SID, we must look at the similarities with the CWT Frazier [1999]. The general formula of the CWT is expressed by the integral

$$F(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\ \overline{\psi\left(\frac{t-b}{a}\right)}\ dt, \tag{8.10}$$

where $F(a,b)$ is continuous-time wavelet transform of square-integrable function $f(t) \in \ell^2(\mathbb{R})$, $\psi(t)$ is (mother) wavelet function and $a, b \in \mathbb{R}$, $a > 0$ are scale and shift, respectively. Comparing this formula with the integral

$$\int_0^T f(t)\phi(t)dt \tag{8.11}$$

used within the modulating function method, we can state some relations between wavelets and modulating functions. First and the most important remark is that the modulating function $\phi(t)$ corresponds to the complex conjugate of the wavelet $\overline{\psi(t)}$ and not to the wavelet $\psi(t)$ itself. Next, the shifts and scales of the basic wavelet could be used as modulating functions as well. The requirements imposed on modulating functions must be still satisfied for $\overline{\psi(t)}$. An inner product with a complex conjugate is a convolution with a conjugate reflection (4.8), thus the wavelet transform (8.10) performs a filtering and the symmetry of wavelet function as a modulating function is not required anymore. Now it can seem that the condition "to be a modulating function" is stronger than the condition "to be a wavelet", but it is not true and it is not weaker condition either. Both wavelets and modulating functions are special

types of functions themselves, there are modulating functions which are not wavelets and there are wavelets which are not modulating functions. Note that one of the first notes on similarities of modulating functions and wavelets was mentioned in Preisig [2001], where the relation between Maletinsky's spline modulating functions and B-spline wavelets is discussed. Next, Mukhopadhyay et al. [2010] handles the spline wavelets for SID problem.

There is one more advanced feature of the CWT which is possible to incorporate into the modulating functions method to simplify the whole procedure. It is the Mallat's algorithm (Frazier [1999]) enabling fast computation of the CWT (sometimes called fast wavelet transform). The Mallat's algorithm enables to compute the transform at all levels with use of basic wavelets only. Wavelet transform coefficients at higher levels are then computed based on the knowledge of the coefficients at one level lower. In connection with necessity to produce several algebraic equation to solve the identification problem, the Mallat's algorithm simplifies only that computations which use the same basic wavelet.

As we already know from Section 4.5 and Section 4.5.1, continuous-time wavelets usually have explicit expression, what is an advantage for computing the integral (8.10). There are, however, other troubles with application of the modulating function method with wavelets. Foremost, the essential problem of the pure modulating functions method is that both input and output must be also expressed explicitly. Another issue is that continuous-time wavelets have usually infinite support, so they do not comply one of the key requirements and therefore they can not be used as a modulating function in a classical manner. Nevertheless, as the real data are always sampled no matter the system description is continuous or discrete, the integrals within the modulating functions method can be computed through a numerical integration. Therefore, the continuous-time wavelets have to be approximated by their discrete expression (as mentioned in Section 4.5.1).

Finally, the whole identification problem becomes discrete. It however can not be said that the CWT is replaced by the DWT. Realize that it is not necessary to employ the wavelet transform in its essence, but only the wavelet functions with their specific properties as advantages are to be utilized as modulating functions. Therefore, there is no necessity of keeping the rule of even shifts of wavelets and we can shift wavelet function just by 1 sample. In order to do that, let us look at the wavelets from the linear systems point of view. Considering a linear dynamical filter to have the same filtering properties as particular wavelet function, this filter has to generate the wavelet function as its impulse response. Next, considering a discrete-time approximation of the wavelet (when necessary), the resulting filter is a simple FIR filter and can be computed as

$$FIR(q) = \sum_{j=0}^{L-1} q^{-j}\psi(j),$$

where $L-1$ is the order of the filter (what corresponds to the support L of the wavelet $\psi(t)$). The example and comparison of such LTI filters is depicted in Figure 8.1. The

advantage of approach via FIR filters is definitely a possibility to decrease mutual shift of wavelets, thus reducing necessary length of measured data, and gaining the possibility to apply a recursive identification with each new incoming data point, although at the expense of lost of orthogonality. On the other hand, the principle of DWT can be preserved when applying wavelets in a way that the rules of DWT are complied. Then, comparing the formulae (4.13) with the discrete approximation to modulating functions method we again obtain that the modulating function should be a complex conjugate of wavelet function, even in a discrete approach.

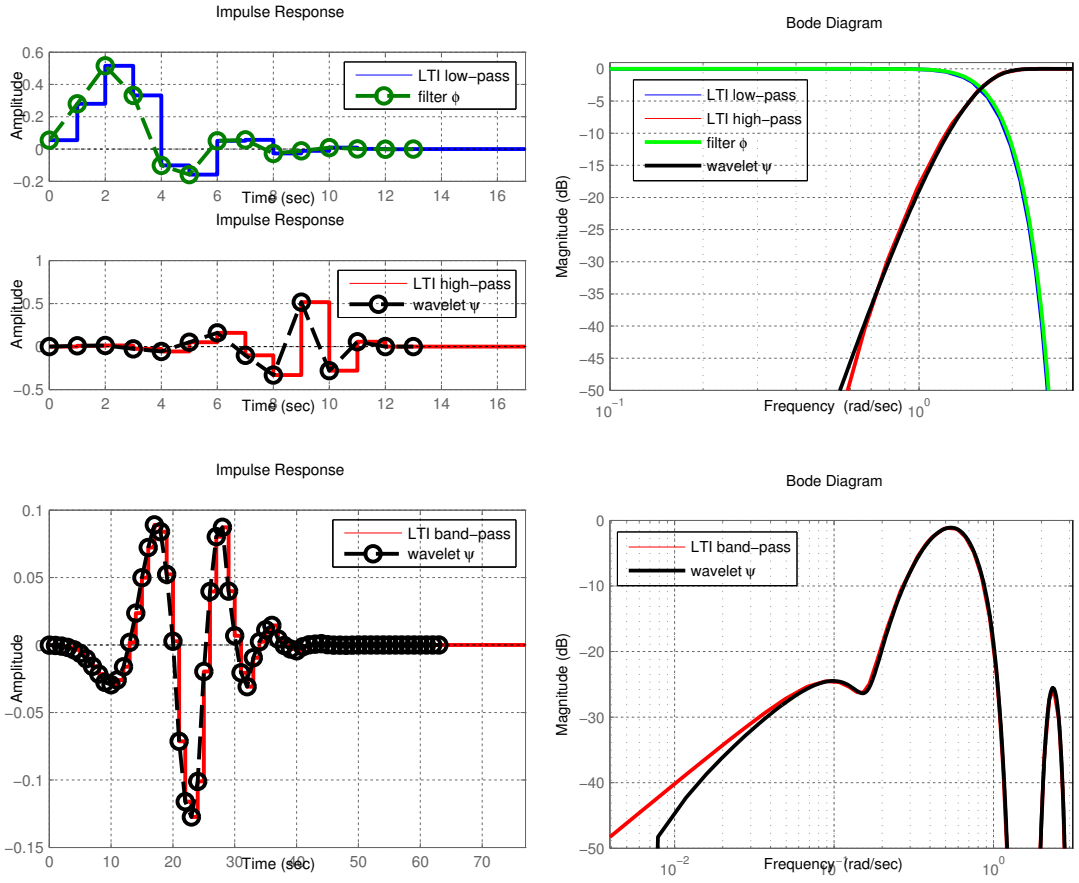


Figure 8.1: Comparison of FIR filters and appropriate wavelets in both time and frequency domains.
Left figures depict time-domain, right figures depict frequency domain. Upper figures concerns both wavelet and filter db7 at the 1^{st} level and their approximation by FIR filters and depict a division of the whole frequency domain into halves by low-pass and high-pass filters. The lower figures concerns the wavelet db4 at the 3^{rd} level, what is band-pass filter.

Realize a very important fact that the equation (6.20) for the variance of the estimate still holds when orthogonal wavelets are used. There are different explanations for it: i) Parseval's theorem holds in general and does not depend on the particular shift of the wavelet functions. ii) Equation (6.20) itself does not depend on a shift. iii) Using of any shift of wavelet functions does not affect their frequency characteristics, but the length of data used for the identification only. Shorter shift of wavelet functions causes just a duplicity of (time-frequency) information borne in wavelet coefficients, but an amount of that information is the same as while using correct shift. Longer shift (what is not a case of FIR filters) means that some data points are not subject of wavelet transform, thus the variance can be worse than when using correct shift. However, this difference is not caused by filters themselves, but by less data points used for an identification and hence is covered by number of data points N. Moreover, a very large shifting can cause omitting some data, thus change of spectra $\Phi_u(m)$ of data used for the identification.

Another comment on the error of estimation follows. Consider for a while that one have explicit expressions for both input and output signals and that orthogonal (thus discrete-time) wavelets are to be used. A numerical integration has to be employed due to wavelets. Then when computing an error of the estimate, we should not forget to count on the error caused by the numerical integration itself. When using continuous-time non-orthogonal wavelets, the integration can be done analytically, thus exactly. When using both discrete-time data and discrete-time wavelets, talking about the error of integration has no sense, however, another error caused by an approximation of derivatives of wavelets in discrete-time arises. These errors are inherently associated to the method itself and can be suppressed by a suitable choice of numerical approach of computation of either integral or derivative, e.g. sufficiently fast sampling of data together with using wavelets of high orders.

8.2.2 Choice of wavelet modulating function

This section is analogy to the Section 8.1.1 and treats the selection of wavelets as modulating functions only. Indeed, all the properties previously mentioned in Section 8.1.1 still hold, so mainly additional possibilities and points of view on wavelets are discussed here.

1. **Type:** In spite of the number of wavelet families, when considering some particular wavelet function, different wavelet modulating functions can be obtained by their shifting or scaling. The scaling means their stretching or shrinking together with their normalization. Next operation modifying the basic wavelet and resulting into the different wavelet modulating function is the making the wavelet filters corresponding to the higher level analysis.

   There is yet another method to compute discrete-time wavelet functions for higher analysis levels, which results completely from the DWT theory. It determines the computation of partial filters summed in Table 4.1 in time domain. As these

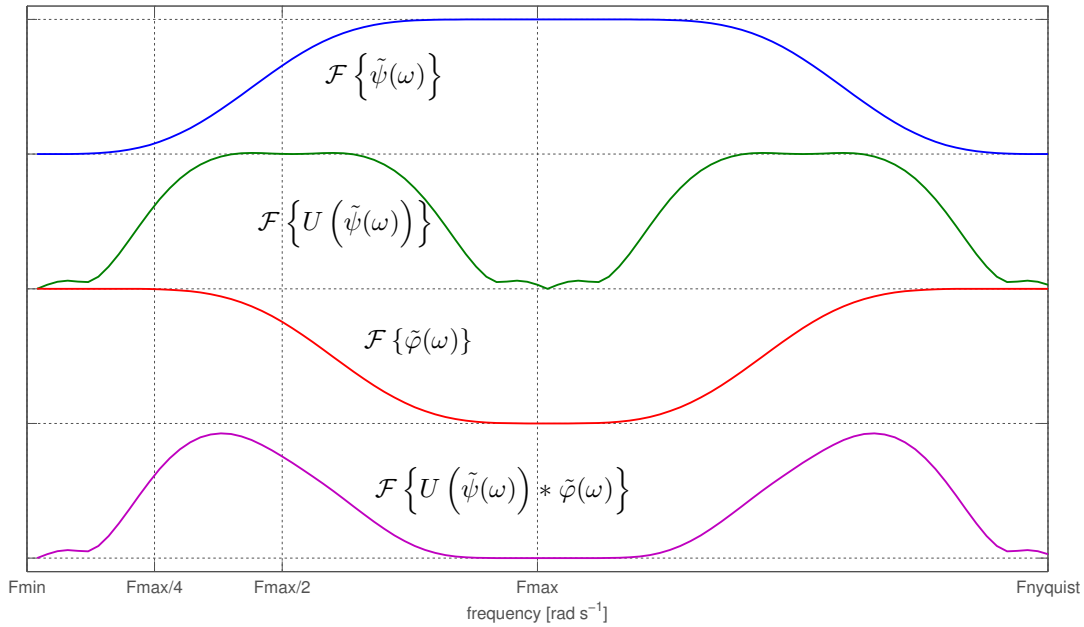


Figure 8.2: A frequency-domain demonstration of computation of discrete wavelet functions in time domain.

filters were marked by $\hat{w}_j,\ j = 1, \ldots, p+1$ (Section 6.3), we can write

$$\begin{aligned} w_1(k) &= \tilde{\psi}(k), \\ w_2(k) &= U\left(\tilde{\psi}(k)\right) * \tilde{\varphi}(k) = U\left(w_1(k)\right) * \tilde{\varphi}(k), \\ w_3(k) &= U\left(w_2(k)\right) * U\left(\tilde{\varphi}(k)\right) * \tilde{\varphi}(k), \\ &\vdots \end{aligned}$$

what consequently yields the recursive formulae

$$w_i(k) = U\left(w_{i-1}(k)\right) * U^{i-2}\left(\tilde{\varphi}(k)\right) * U^{i-3}\left(\tilde{\varphi}(k)\right) * \cdots * \tilde{\varphi}(k), \tag{8.12}$$

where U^i stands for i times repeated operator $U(\bullet)$.

The procedure can be very simply understood in a frequency domain and is shown in the Figure 8.2. The shapes of frequency characteristics of basic filters $\tilde{\varphi}, \tilde{\psi}$ are already known from Section 4.4. The second line in the Figure 8.2 shows how does the upsampling operator $U(\bullet)$ transform the frequency properties of the function: it shrinks the spectra of the original function and extend it periodically to the full range. Afterward, the convolution with father wavelet is performed, what is just multiplying in frequency with appropriate low-pass filter, see the last line in the Figure 8.2. The result is a new wavelet function, which represents the

analysis at the 2^{nd} level. Note that its frequency characteristics is the same as shown in the Figure 4.2.

All these operations preserve the properties required on the function to be a modulating function and can be combined arbitrarily.

2. **Sampling:** The sampling corresponds mainly to the ability of the wavelet modulating function to get a particular frequency information from the data. There is a big difference between using discrete-time wavelets and using discrete approximation of continuous-time wavelets.

   When employing discrete-time wavelets, their length is strictly given by the type of wavelet and the sampling is given by a measuring, so the possibilities of frequency domain division are predetermined as described in Section 4.4. The only choice for required frequency range selection is then the procedure described above in the point 1.

   On the other hand, the discrete approximation of continuous-time wavelets provides the user with an unique choice of sampling time for wavelets themselves. Together with the combination of operations from the previous point, the user can ensure arbitrary frequency localisation of the wavelet modulating functions for the specific application.

3. **Independence in time or frequency:** This property has been partially discussed within the basic properties of wavelets in Section 4.5, therefore we only emphasize the important remarks here.

   The discrete-time wavelets are usually orthogonal among their scales and even shifts, what plays a very important role for independence of the finally obtained algebraic equations. However the orthogonality in time does not guarantee the independence of the algebraic equations itself. Moreover, the mutual shifts of measured data corresponding to the shifted wavelets should not be correlated, what can be managed by the sufficient excitation of the system. It finally ensures the independence of the wavelet transform coefficients, or in this case, coefficients of algebraic equations. Due to a finite support, these wavelets overlap in frequency domain and are not orthogonal there at all, nevertheless, they are practically exclusive.

   Note that there is no rule about the shift of modulating functions, but classical shift $2^i k$ for the i^{th} level wavelet analysis assures the qualities of wavelet transform as well. Nevertheless, utilizing of wavelets as modulating functions does not oblige the user to employ the wavelet transform itself, thus it is not necessary to comply those even shifts. The trade-off has to be done about the shifts: The shift guaranteeing an orthogonality is longer, so more data is needed. On the contrary, simple shift by 1 point needs less data, but the coefficients could bear a duplicate information. This obviously holds across the scaling. Realize yet that for identification with focus on low frequencies, the appropriate wavelets have long

support. Joining this fact with several shifts of a basic filter (for more algebraic equations) results in necessity of long data as well.

Concerning the continuous-time wavelets as modulating functions, the situation is the opposite. The wavelets (or rather their discrete approximations) are not orthogonal, but can have an excellent localisation in frequency. Therefore the coefficients in algebraic equations can not be fully independent and it is a must to generate more equations than the number of parameters. On the other hand, user-defined frequency localisation of these wavelets enables very precise frequency localisation of the identified system.

9 Conclusions

9.1 Results

The thesis presented a new approaches to utilization of the wavelet transform in the field of system identification. Since the wavelet transform as a mathematical tool serves mainly for signal analysis both in time and frequency domains, the algorithm introduced in the thesis represents a natural way of interconnection of discrete wavelet theory and theory of system identification. This approach is based mainly on favourable properties of wavelet basis functions and brings several advantages:

1. The set of wavelet basis functions at all possible levels (i.e. all scales of one particular couple of wavelet basis functions - father and mother wavelets) forms a set of filters. This set then covers the whole frequency range determined by the properties of analysed signal, more specifically, by its length and a sampling time.

2. Due to theoretical restrictions, orthogonal wavelets have compact support thus simple structure. Consequently, all convolutions are exactly computable, thus no information carried in signal is lost.

3. Some wavelet filters are orthogonal in time and complementary in frequency domains, therefore each filter extracts the specific portion of information from the signals without any duplicity. This fact also contributes to numerical conditioning of the identification algorithm.

4. Moreover, there are several kinds of wavelet basis functions (wavelet families) with different time or frequency properties. Most of them satisfy the necessary conditions given by the wavelet theory, however, there are also some kinds of wavelet functions, which do not. These exceptions are covered by theory of wavelet frames, which is generalization of the wavelet theory and is built on the theory of Riesz's basis.

5. Once implemented, this method is quite generic and intuitive, while the design of appropriate linear filters could be quite time consuming. In addition, this provides us with a big advantage in real problems, where the frequency characteristics of the system to be identified is not known a priori. The satisfactory results could be acquired by tuning of weights only, which corresponds to some knowledge of the system. On the other side, the implementation itself requires deeper understanding to wavelet theory.

At the beginning, the thesis discusses the possibilities of incorporation of wavelet transform into the system identification in the form of a comprehensive study. Then, a general concept of the incorporation of wavelets was introduced. At first, the method was derived for simple SISO systems as well as its asymptotic properties were discussed. Regarding the asymptotic properties of the proposed methods, the results were derived based on properties of the PEM for identification of the ARX model. Further on in the thesis, the proposed concept was extended on firstly for MISO and then for MIMO systems. The algorithms for all parts were implemented and their performance was demonstrated on case studies at the end of appropriate chapters. Finally, a detailed analysis of the utilizing the wavelets as modulating functions were elaborated.

Although the wavelets are used, in principle it still is proper selection and (pre)filtering of data with all its pros and cons. There are mainly 2 ways where the methods from the thesis can be employed with advantages:

a) A sufficiently accurate model is required, well describing the system's behaviour at particular frequency ranges.

b) The only submodel is identified (the lowest order model as possible) which takes into account behaviour on required frequency ranges. It includes ability to identify slow or fast subsystems of singularly perturbed system as well as to do model reduction for identification for control.

Indeed, both ways can be linked together.

9.2 Fulfillment of the objectives

Here a short note on fulfilment of the aims from Chapter 2 is provided.

√ To perform a comprehensive survey of the methods of exploiting the wavelet transform for system identification. → This objective was completed and described in Chapter 3.

√ To find and describe a suitable way of incorporation of wavelet transform into the problem of general single-input single-output linear system identification. Analyse the method and demonstrate it on a suitable example. → This objective was satisfied by development and description of the algorithm, its implementation and demonstration on an example. All is stated within the Chapter 6. This is

the main part of the author's publications Váňa and Preisig [2012]; Váňa et al. [2011].

√ To extend the method to multivariable systems. Analyse the method and demonstrate it on a suitable example. → This objective was handled in several points of view on multivariable systems description. The ways of incorporation of wavelet transform into multivariable system identification was described and demonstrated within the Chapter 7. This objective was also partially described in Váňa and Preisig [2012]; Váňa et al. [2011], partially since at that time, the work on the objective was still in progress.

√ To investigate and find the utilization of wavelet transform within the continuous-time linear system identification. The discussion on implementation issues must be included. → This objective was satisfied by the Chapter 8, where the modulating function method primarily designed for continuous-time system identification were shown to be a great alternative for utilization of wavelets within system identification.

All the algorithms implemented within this work are available on the enclosed CD. Moreover, the reader can find there those examples used within case studies, which are not adherent to any coordination with some commercial company.

9.3 Concluding remarks

The thesis builds on the existing literature results and interprets them differently besides clarifying properties and links between methods and approaches. To be more specific, the thesis provides several points of view on wavelets, what enables the reader to understand both the wavelets and mutual consequences between wavelets and system identification more deeply. The most of relations of theories of wavelet transform and of system identification were described in a general manner, therefore they can be simply applied in different fields.

Except of the algorithms described in the thesis, the interconnection of theories of both wavelet transform and system identification is the main contribution of the thesis. Nowadays, there are lot of distinct highly professional tools, which are well-known and widely used by people from some particular branch only. It is therefore very important not only to develop new methods, but also to look for suitable methods across different scientific fields. There is a strong reason for doing it even in situations where no new results can be obtained, since using of tools which are "new" in particular field can always show new analogies and links.

Author's International Reviewed Journals

Z. Váňa and H.A. Preisig. System identification in frequency domain using wavelets: Conceptual remarks. *Systems and Control Letters*, 61 (10):1041–1051, 2012. ISSN 0167-6911. doi: http://dx.doi.org/10.1016/j.sysconle.2012.07.004. URL http://www.sciencedirect.com/science/article/pii/S0167691112001375. (co-authorship: 95%).

Z. Váňa, J. Cigler, J. Široký, E. Žáčeková, and L. Ferkl. Model-based energy efficient control applied to an office building. *Journal of Process Control*, 2014. ISSN 0959-1524. accepted on 30th January 2014.

J. Cigler, S. Prívara, Z. Váňa, E. Žáčeková, and L. Ferkl. Optimization of predicted mean vote index within model predictive control framework: Computationally tractable solution. *Energy and Buildings*, 52(0):39 – 49, 2012. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.05.022. URL http://www.sciencedirect.com/science/article/pii/S0378778812002770. (co-authorship: 15%).

S. Prívara, J. Cigler, Z. Váňa, and L. Ferkl. Incorporation of system steady state properties into subspace identification algorithm. *International Journal of Modelling, Identification and Control*, 16(2):159–167, 2012a. doi: 10.1504/IJMIC.2012.047123. URL http://www.inderscience.com/info/inarticle.php?artid=47123. (co-authorship: 30%).

S. Prívara, Z. Váňa, E. Žáčeková, and J. Cigler. Building modeling: Selection of the most appropriate model for predictive control. *Energy and Buildings*, 55:341–350, December 2012b. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.08.040. URL http://www.sciencedirect.com/science/article/pii/S0378778812004446?v=s5. (co-authorship: 30%).

S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, C. Sagerschnig, and E. Žáčeková. Building modeling as a crucial part for building predictive control. *Energy and Buildings*, 56:8–22, January 2013a. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.10.024. URL http://www.sciencedirect.com/science/article/pii/S0378778812005336?v=s5. (co-authorship: 15%).

S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, and E. Žáčeková. Use of partial least squares within the control relevant identification for buildings. *Control Engineering Practice*, 21, January 2013b. ISSN 0967-0661. doi: http://dx.doi.org/10.

1016/j.conengprac.2012.09.017. URL http://www.sciencedirect.com/science/article/pii/S0967066112002006. (co-authorship: 15%).

Author's International Conferences

Z. Váňa, S. Prívara, J. Cigler, and H.A. Preisig. System identification using wavelet analysis. In *21st European Symposium on Computer Aided Process Engineering (ESCAPE)*, volume 29 of *Computer Aided Chemical Engineering*, pages 763–767. Elsevier, Porto Carras, Greece, 2011. doi: DOI:10.1016/B978-0-444-53711-9.50153-X. (co-authorship: 50%).

O. Bruna and Z. Váňa. Parameters identification of a chemical tank: A case study. In *20th Mediterranean Conference on Control and Automation (MED)*, pages 241–246, Barcelona, Spain, 2012. IEEE. ISBN 978-1-4673-2530-1. doi: 10.1109/MED.2012.6265854. (co-authorship: 20%).

J. Cigler, S. Prívara, Z. Váňa, D. Komárková, and M. Šebek. Optimization of predicted mean vote thermal comfort index within model predictive control framework. In *51th IEEE Conference on Decision and Control (CDC)*, Maui, Hawaii, US, December 2012a. IEEE. ISBN 978-1-4673-2064-1. doi: 10.1109/CDC.2012.6427051. (co-authorship: 15%).

J. Cigler, S. Prívara, Z. Váňa, E. Žáčeková, and L. Ferkl. On predicted mean vote optimization in building climate control. In *20th Mediterranean Conference on Control and Automation (MED)*, pages 1518–1523, Barcelona, Spain, 2012b. IEEE. ISBN 978-1-4673-2530-1. doi: 10.1109/MED.2012.6265854. (co-authorship: 10%).

S. Prívara, J. Cigler, Z. Váňa, L. Ferkl, and M. Šebek. Subspace identification of poorly excited industrial systems. In *49th IEEE Conference on Decision and Control (CDC)*, pages 4405–4410, Atlanta, US, Dec. 2010. doi: 10.1109/CDC.2010.5717585. (co-authorship: 30%).

S. Prívara, Z. Váňa, J. Cigler, and J. Komárek. Role of mpc in building climate control. In *21st European Symposium on Computer Aided Process Engineering (ESCAPE)*, volume 29 of *Computer Aided Chemical Engineering*, pages 728–732. Elsevier, 2011a. doi: DOI:10.1016/B978-0-444-53711-9.50146-2. (co-authorship: 30%).

S. Prívara, Z. Váňa, D. Gyalistras, J. Cigler, C. Sagerschnig, M. Morari, and L. Ferkl. Modeling and identification of a large multi-zone office building. In *2011 IEEE International Conference on Control Applications (CCA)*, pages 55–60, Denver, US, 2011b. IEEE. doi: 10.1109/CCA.2011.6044402. (co-authorship: 30%).

S. Prívara, Z. Váňa, J. Cigler, and L. Ferkl. Predictive control oriented subspace identification based on building energy simulation tools. In *20th Mediterranean Conference on Control and Automation (MED)*, pages 1290–1295, Barcelona, Spain, 2012. IEEE. ISBN 978-1-4673-2530-1. doi: 10.1109/MED.2012.6265817. (co-authorship: 30%).

C. Sagerschnig, D. Gyalistras, A. Seerig, S. Prívara, J. Cigler, and Z. Váňa. Co-simulation for building controller development: the case study of a modern office building. In *CleanTech for Sustainable Buildings – From Nano to Urban Scale*, pages 955–960, Lausanne, Switzerland, 2011. Solar Energy nad Building Physics Laboratory of the Swiss Federal Institute of Technology Lausanne. ISBN 978-2-8399-0906-8. (co-authorship: 20%).

Z. Váňa, J. Kubeček, and L. Ferkl. Notes on finding black-box model of a large building. In *2010 IEEE International Conference on Control Applications (CCA)*, pages 1017–1022. IEEE, 2010. (co-authorship: 40%).

Z. Váňa, S. Prívara, E. Žáčeková, and J. Cigler. Building semi-physical modeling: On selection of the model complexity. In *Proceedings of the European Control Conference (ECC)*, pages 3955–3960, Zurich, Switzerland, 2013. IEEE. ISBN 978-3-9524173-4-8. (co-authorship: 35%).

E. Žáčeková and Z. Váňa. Identification and model selection of building models. In *20th Mediterranean Conference on Control and Automation (MED)*, pages 1512–1517, Barcelona, Spain, 2012. IEEE. ISBN 978-1-4673-2530-1. doi: 10.1109/MED.2012.6265817. (co-authorship: 30%).

E. Žáčeková, S. Prívara, and Z. Váňa. Model predictive control relevant identification using partial least squares for building modeling. In *Australian Control Conference (AUCC), 2011*, pages 422–427. IEEE, 2011. (co-authorship: 25%).

E. Žáčeková, S. Prívara, Z. Váňa, and J. Cigler. Building modeling: on selection of the model complexity for predictive control. In *23rd European Symposium on Computer Aided Process Engineering (ESCAPE)*, Computer Aided Chemical Engineering, pages 205–210. Elsevier, Lappeenranta, Finland, 2013a. ISBN 978-0-444-63234-0. (co-authorship: 25%).

E. Žáčeková, S. Prívara, Z. Váňa, J. Cigler, and L. Ferkl. Dual control approach for zone model predictive control. In *Proceedings of the European Control Conference (ECC)*, pages 1398–1403, Zurich, Switzerland, 2013b. IEEE. ISBN 978-3-9524173-4-8. (co-authorship: 10%).

Author's Domestic Conferences

J. Kubeček, Z. Váňa, J. Cigler, and S. Prívara. Identification and predictive control of large multiple input multiple output systems. In *11th Poster Conference*, May 2010.

Z. Váňa, J. Cigler, and S. Prívara. Modeling for control of a large office building. In *11th Poster Conference*, May 2011a.

Z. Váňa, J. Cigler, and S. Prívara. Wavelet filters in system identification. In *11th Poster Conference*, May 2011b.

References

H. Adeli and X. Jiang. Dynamic fuzzy wavelet neural network model for structural system identification. *Journal of Structural Engineering*, 132(1):102–111, 2006. ISSN 0733-9445. doi: 10.1061/(ASCE)0733-9445(2006)132:1(102).

N. Ahuja, S. Lertrattanapanich, and N. K. Bose. Properties determining choice of mother wavelet. *Vision, Image and Signal Processing, IEE Proceedings* -, 152(5): 659–664, 2005. ISSN 1350-245X. doi: 10.1049/ip-vis:20045034.

P. J. Antsaklis and A. N. Michel. *Linear systems*. Birkhauser Boston, 1997.

S. A. Billings and H. L. Wei. A new class of wavelet networks for nonlinear system identification. *IEEE Transactions on Neural Networks*, 16(4):862–874, 2005. ISSN 1045-9227. doi: 10.1109/TNN.2005.849842.

M. Boltezar and J. Slavic. Enhancements to the continuous wavelet transform for damping identifications on short signals. *Mechanical Systems and Signal Processing*, 18(5):1065–1076, 2004. ISSN 0888-3270. doi: 10.1016/j.ymssp.2004.01.004. URL http://www.sciencedirect.com/science/article/pii/S0888327004000093.

J. F. Carrier and G. Stephanopoulos. Wavelet-based modulation in control-relevant process identification. *AIChE Journal*, 44(2):341–360, 1998. ISSN 1547-5905. doi: 10.1002/aic.690440212. URL http://onlinelibrary.wiley.com/doi/10.1002/aic.690440212/abstract.

C. C. Chang and Y. Shi. Identification of time-varying hysteretic structures using wavelet multiresolution analysis. *International Journal of Non-Linear Mechanics*, 45(1):21–34, 2010. ISSN 0020-7462. doi: 10.1016/j.ijnonlinmec.2009.08.009. URL http://www.sciencedirect.com/science/article/pii/S0020746209001644.

C. K. Chui and J. Z. Wang. A cardinal spline approach to wavelets. *Proceedings of the American Mathematical Society*, 113(3):785–793, 1991. ISSN 0002-9939, 1088-6826. doi: 10.1090/S0002-9939-1991-1077784-X. URL http://www.ams.org/proc/1991-113-03/S0002-9939-1991-1077784-X/.

C. K. Chui and J. Z. Wang. On compactly supported spline wavelets and a duality principle. *Transactions of the American Mathematical Society*, 330(2):903, April 1992. ISSN 00029947. doi: 10.2307/2153941. URL

http://www.jstor.org/discover/10.2307/2153941?uid=3737856&uid=2129&uid=2&uid=70&uid=4&sid=21102515835077.

I. Daubechies et al. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

Y. Dorfan, A. Feuer, and B. Porat. Modeling and identification of lptv systems by wavelets. *Signal processing*, 84(8):1285–1297, 2004.

M. Doroslovacki, H. Fan, and L. Yao. Wavelet-based identification of linear discrete-time systems: Robustness issue. *Automatica*, 34(12):1637–1640, 1998. ISSN 0005-1098. doi: 10.1016/S0005-1098(98)80020-4. URL http://www.sciencedirect.com/science/article/pii/S0005109898800204.

Y. Engel, S. Mannor, and R. Meir. The kernel recursive least-squares algorithm. *IEEE Transactions on Signal Processing*, 52(8):2275–2285, 2004. ISSN 1053-587X. doi: 10.1109/TSP.2004.830985.

S. Erlicher and P. Argoul. Modal identification of linear non-proportionally damped systems by wavelet transform. *Mechanical Systems and Signal Processing*, 21(3): 1386–1421, 2007. ISSN 0888-3270. doi: 10.1016/j.ymssp.2006.03.010. URL http://www.sciencedirect.com/science/article/pii/S0888327006000719.

M. Frazier. *An Introduction to Wavelets Through Linear Algebra*. Springer, June 1999. ISBN 9780387986395.

D. L. Fugal. *Conceptual wavelets in digital signal processing: an in-depth, practical approach for the non-mathematician*. Space & Signals Technical Pub., San Diego, Calif., 2009. ISBN 9780982199459 0982199457.

R. X. Gao and R. Yan. Wavelet packet transform. In *Wavelets*, pages 69–81. Springer, 2011.

H. Garnier, M. Mensler, and A. Richard. Continuous-time model identification from sampled data: implementation issues and performance evaluation. *International Journal of Control*, 76(13):1337–1357, 2003.

M. Gevers. A personal view of the development of system identification: A 30-year journey through an exciting field. *Control Systems, IEEE*, 26(6):93–105, 2006.

R. Ghanem and F. Romeo. A wavelet-based approach for the identification of linear time-varying dynamical systems. *Journal of Sound and Vibration*, 234(4): 555–576, 1999. ISSN 0022-460X. doi: 10.1006/jsvi.1999.2752. URL http://www.sciencedirect.com/science/article/pii/S0022460X99927529.

R. Ghanem and F. Romeo. A wavelet-based approach for model and parameter identification of non-linear systems. *International Journal of Non-Linear Mechanics*, 36(5):835–859, 2001. ISSN 0020-7462. doi: 10.1016/S0020-7462(00)00050-0. URL http://www.sciencedirect.com/science/article/pii/S0020746200000500.

C. E. Heil and D. F. Walnut. Continuous and discrete wavelet transforms. *SIAM Review*, 31(4):628–666, December 1989. ISSN 0036-1445, 1095-7200. doi: 10.1137/1031129. URL http://epubs.siam.org/doi/abs/10.1137/1031129.

K. C. Ho and S. D. Blunt. Adaptive sparse system identification using wavelets. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 49 (10):656–667, 2003. ISSN 1057-7130. doi: 10.1109/TCSII.2002.807263.

C. S. Huang and W. C. Su. Identification of modal parameters of a time invariant linear system by continuous wavelet transformation. *Mechanical Systems and Signal Processing*, 21(4):1642–1664, 2007. ISSN 0888-3270. doi: 10.1016/j.ymssp.2006.07.011. URL http://www.sciencedirect.com/science/article/pii/S0888327006001427.

D. J. Joo. Damage detection and system identification using a wavelet energy based approach. *PhD thesis*, 2012. URL http://academiccommons.columbia.edu/catalog/ac:152497.

T. Kijewski and A. Kareem. Wavelet transforms for system identification in civil engineering. *Computer-Aided Civil and Infrastructure Engineering*, 18(5):339–355, 2003. ISSN 1467-8667. doi: 10.1111/1467-8667.t01-1-00312. URL http://onlinelibrary.wiley.com/doi/10.1111/1467-8667.t01-1-00312/abstract.

Y. Kinoshita and Y. Ohta. Continuous-time system identification using compactly-supported filter kernels generated from laguerre basis functions. In *2010 49th IEEE Conference on Decision and Control (CDC)*, pages 4461–4466, 2010. doi: 10.1109/CDC.2010.5717196.

P. V. Kokotovic and H. K. Khalil, editors. *Singular perturbations in systems and control*. IEEE Press, Piscataway, NJ, USA, 1986. ISBN 0-879-42205-X.

D. Kolzow. Wavelets; a tutorial and a bibliography. *Rendiconti dell'Istituto di matematica dell'Universita di Trieste*, page 49, 1994. ISSN 0049-4704. URL http://www.openstarts.units.it/dspace/handle/10077/4624.

K. A. Kosanovich, A. R. Moser, and M. J. Piovoso. Poisson wavelets applied to model identification. *Journal of Process Control*, 5(4):225–234, 1995. ISSN 0959-1524. doi: 10.1016/0959-1524(95)00012-F. URL http://www.sciencedirect.com/science/article/pii/095915249500012F.

I. A. Kougioumtzoglou and P. D. Spanos. An identification approach for linear and nonlinear time-variant structural systems via harmonic wavelets. *Mechanical Systems and Signal Processing*, 37(1–2):338–352, 2013. ISSN 0888-3270. doi: 10.1016/j.ymssp.2013.01.011. URL http://www.sciencedirect.com/science/article/pii/S0888327013000319.

J. Lardies, M. N. Ta, and M. Berthillier. Modal parameter estimation based on the wavelet transform of output data. *Archive of Applied Mechanics*, 73(9-10):718–733, April 2004. ISSN 0939-1533, 1432-0681. doi: 10.1007/s00419-004-0329-6. URL http://link.springer.com/article/10.1007/s00419-004-0329-6.

J. Li and J. H. Liu. Identification and control of dynamic systems based on least squares wavelet vector machines. In J. Wang, Z. Yi, J. M. Zurada, B. L. Lu, and H. Yin, editors, *Advances in Neural Networks - ISNN 2006*, number 3972 in Lecture Notes in Computer Science, pages 934–942. Springer Berlin Heidelberg, January 2006. ISBN 978-3-540-34437-7, 978-3-540-34438-4. URL http://link.springer.com/chapter/10.1007/11760023_138.

Y. Li, H. L. Wei, S. A. Billings, and P. G. Sarrigiannis. Time-varying model identification for time–frequency feature extraction from eeg data. *Journal of neuroscience methods*, 196(1):151–158, 2011.

L. Ljung. *System identification: Theory for the user*. 1999.

L. Ljung. Prediction error estimation methods. *Circuits, Systems and Signal Processing*, 21(1):11–21, January 2002. ISSN 0278-081X, 1531-5878. doi: 10.1007/BF01211648. URL http://link.springer.com/article/10.1007/BF01211648.

S. Ljung and L. Ljung. Error propagation properties of recursive least-squares adaptation algorithms. *Automatica*, 21(2):157–167, 1985. ISSN 0005-1098. doi: 10.1016/0005-1098(85)90110-4. URL http://www.sciencedirect.com/science/article/pii/0005109885901104.

R. W. P. Luk and R. I. Damper. Non-parametric linear time-invariant system identification by discrete wavelet transforms. *Digital Signal Processing*, 16(3):303–319, 2006. ISSN 1051-2004. doi: 10.1016/j.dsp.2005.11.004. URL http://www.sciencedirect.com/science/article/pii/S1051200405001685.

S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, September 1999. ISBN 9780080520834.

H. P. Mohallem and R. H. G. Kawakami. Wavelet-packet identification of dynamic systems in frequency subbands. *Signal Processing*, 86(8):2001–2008, 2006. ISSN 0165-1684. doi: 10.1016/j.sigpro.2005.09.021. URL http://www.sciencedirect.com/science/article/pii/S0165168405003543.

S. Mukhopadhyay and A. P. Tiwari. Consistent output estimate with wavelets: An alternative solution of least squares minimization problem for identification of the LZC system of a large PHWR. *Annals of Nuclear Energy*, 37(7):974–984, 2010. ISSN 0306-4549. doi: 10.1016/j.anucene.2010.03.006. URL http://www.sciencedirect.com/science/article/pii/S0306454910000964.

S. Mukhopadhyay, U. Mahapatra, A. K. Tangirala, and A. P. Tiwari. Spline wavelets for system identification. *Proceedings of DYCOPS 2010*, 2010.

R. Polikar. The story of wavelets. *Physics and modern topics in mechanical and electrical engineering*, pages 192–197, 1999.

H. A. Preisig. Using wavelets in process identification: A new link to the state space. *Computer Aided Chemical Engineering*, 9:261–266, 2001.

H. A. Preisig. Parameter estimation using multi-wavelets. In S. Pierucci and G. Buzzi Ferraris, editor, *Computer Aided Chemical Engineering*, volume Volume 28, pages 367–372. Elsevier, 2010. ISBN 1570-7946. URL http://www.sciencedirect.com/science/article/pii/S1570794610280628.

H. A. Preisig and D. W. T. Rippin. (a) theory and application of the modulating function method—I. review and theory of the method and theory of the spline-type modulating functions. *Computers & Chemical Engineering*, 17(1):1–16, 1993a. ISSN 0098-1354. doi: 10.1016/0098-1354(93)80001-4. URL http://www.sciencedirect.com/science/article/pii/0098135493800014.

H. A. Preisig and D. W. T. Rippin. (b) theory and application of the modulating function method—II. algebraic representation of maletinsky's spline-type modulating functions. *Computers & Chemical Engineering*, 17(1):17–28, 1993b. ISSN 0098-1354. doi: 10.1016/0098-1354(93)80002-5. URL http://www.sciencedirect.com/science/article/pii/0098135493800025.

H. A. Preisig and D. W. T. Rippin. (c) theory and application of the modulating function method—III. application to industrial process, a well-stirred tank reactor. *Computers & Chemical Engineering*, 17(1):29–39, 1993c. ISSN 0098-1354. doi: 10.1016/0098-1354(93)80003-6. URL http://www.sciencedirect.com/science/article/pii/0098135493800036.

S. Prívara, J. Cigler, Z. Váňa, L. Ferkl, and M. Šebek. Subspace identification of poorly excited industrial systems. In *49th IEEE Conference on Decision and Control (CDC)*, pages 4405–4410, Atlanta, US, Dec. 2010. doi: 10.1109/CDC.2010.5717585.

S. Prívara, J. Široký, L. Ferkl, and J. Cigler. Model predictive control of a building heating system: The first experience. *Energy and Buildings*, 43(2):564–572, 2011.

S. Prívara, Z. Váňa, E. Žáčeková, and J. Cigler. Building modeling: Selection of the most appropriate model for predictive control. *Energy and Buildings*, 55:341–350, December 2012. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.08.040. URL http://www.sciencedirect.com/science/article/pii/S0378778812004446?v=s5.

S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, C. Sagerschnig, and E. Žáčeková. Building modeling as a crucial part for building predictive control. *Energy and Buildings*, 56:8–22, January 2013a. ISSN 0378-7788. doi: 10.1016/j.enbuild.2012.10.024. URL http://www.sciencedirect.com/science/article/pii/S0378778812005336?v=s5.

S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, and E. Žáčeková. Use of partial least squares within the control relevant identification for buildings. *Control Engineering Practice*, 21, January 2013b. ISSN 0967-0661. doi: http://dx.doi.org/10.1016/j.conengprac.2012.09.017. URL http://www.sciencedirect.com/science/article/pii/S0967066112002006.

J. Ramarathnam and A. K. Tangirala. On the use of poisson wavelet transform for system identification. *Journal of Process Control*, 19(1):48–57, 2009. ISSN 0959-1524. doi: 10.1016/j.jprocont.2008.02.001. URL http://www.sciencedirect.com/science/article/pii/S0959152408000267.

G. P. Rao and H. Unbehauen. Identification of continuous-time systems. *Control Theory and Applications, IEE Proceedings* -, 153(2):185–220, 2006. ISSN 1350-2379. doi: 10.1049/ip-cta:20045250.

R. M. Rao, A. S. Bopardikar, and T. Boros. Wavelet transforms: Introduction to theory and applications. *Journal of Electronic Imaging*, 8:478, 1999.

W. Rudin. *Real and complex analysis (3rd)*. New York: McGraw-Hill Inc, 1986.

M. Ruzzene, A. Fasana, L. Garibaldi, and B. Piombo. Natural frequencies and dampings identification using wavelet transform: application to real data. *Mechanical Systems and Signal Processing*, 11(2):207–218, 1997. ISSN 0888-3270. doi: 10.1006/mssp.1996.0078. URL http://www.sciencedirect.com/science/article/pii/S0888327096900783.

Elena Serban. Numerical algorithm for non-linear systems identification based on wavelet transform. In T. Boyanov, S. Dimova, K. Georgiev, and G. Nikolov, editors, *Numerical Methods and Applications*, number 4310 in Lecture Notes in Computer Science, pages 701–708. Springer Berlin Heidelberg, January 2007. ISBN 978-3-540-70940-4, 978-3-540-70942-8. URL http://link.springer.com/chapter/10.1007/978-3-540-70942-8_85.

X. Shan and J. B. Burl. Continuous wavelet based linear time-varying system identification. *Signal Processing*, 91(6):1476–1488, 2011. ISSN 0165-1684. doi: 10.1016/j.sigpro.2010.09.019. URL http://www.sciencedirect.com/science/article/pii/S0165168410003725.

H. L. Shi, D. Y. L. Cai, and Z. L. Qiu. Improved system identification approach using wavelet networks. *Journal of Shanghai University (English Edition)*, 9(2):159–163, April 2005. ISSN 1007-6417, 1863-236X. doi: 10.1007/s11741-005-0070-6. URL http://link.springer.com/article/10.1007/s11741-005-0070-6.

C. Shiguo, Z. Ruanyu, W. Peng, and L. Taihua. Enhance accuracy in pole identification of system by wavelet transform de-noising. *IEEE Transactions on Nuclear Science*, 51(1):250–255, 2004. ISSN 0018-9499. doi: 10.1109/TNS.2004.825098.

M. Shinbrot. On the analysis of linear and nonlinear dynamical systems from transient response data. *Physics and modern topics in mechanical and electrical engineering*, 1954.

J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995. ISSN 0005-1098. doi: 10.1016/0005-1098(95)00120-8. URL http://www.sciencedirect.com/science/article/pii/0005109895001208.

W. J. Staszewski. Identification of non-linear systems using multi-scale ridges and skeletons of the wavelet transform. *Journal of Sound and Vibration*, 214(4):639–658, 1998. ISSN 0022-460X. doi: 10.1006/jsvi.1998.1616. URL http://www.sciencedirect.com/science/article/pii/S0022460X98916169.

G. Strang and V. Strela. Orthogonal multiwavelets with vanishing moments, 1994.

V. Strela and G. Strang. Finite element multiwavelets. In *PROC. MARATEA NATO CONFERENCE, KLUWER*, page 485–496. Kluwer Academic Publishers, 1994.

V. Strela and A. Vogan. Multiwavelets: Theory and applications. Technical report, Massachusetts institute of technology, 1996.

N. Sureshbabu and J. A. Farrell. Wavelet-based system identification for nonlinear control. *IEEE Transactions on Automatic Control*, 44(2):412–417, 1999. ISSN 0018-9286. doi: 10.1109/9.746278.

M. K. Tsatsanis and G. B. Giannakis. Time-varying system identification and model validation using wavelets. *IEEE Transactions on Signal Processing*, 41(12):3512–3523, 1993. ISSN 1053-587X. doi: 10.1109/78.258089.

P. M. J. Van den Hof. *System identification, Lecture notes Wb2403.* 1996.

Z. Váňa and H. A. Preisig. System identification in frequency domain using wavelets: Conceptual remarks. *Systems and Control Letters*, 61(10):1041–1051, 2012. ISSN 0167-6911. doi: http://dx.doi.org/10.1016/j.sysconle.2012.07.004. URL http://www.sciencedirect.com/science/article/pii/S0167691112001375.

Z. Váňa, S. Prívara, J. Cigler, and H. A. Preisig. System identification using wavelet analysis. In *21st European Symposium on Computer Aided Process Engineering (ESCAPE)*, volume 29 of *Computer Aided Chemical Engineering*, pages 763–767. Elsevier, Porto Carras, Greece, 2011. doi: DOI:10.1016/B978-0-444-53711-9.50153-X.

Z. Váňa, J. Cigler, J. Široký, E. Žáčeková, and L. Ferkl. Model-based energy efficient control applied to an office building. *Journal of Process Control*, 2014. ISSN 0959-1524. accepted on 30 January 2014.

J. Široký, F. Oldewurtel, J. Cigler, and S. Prívara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88 (9):3079–3087, 2011.

S. Y. Wang, Y. X. Chang, X. F. Li, and J. G. Zhang. Parameter identification for a class of nonlinear chaotic and hyperchaotic flows. *Nonlinear Analysis: Real World Applications*, 11(1):423–431, 2010. ISSN 1468-1218. doi: 10.1016/j.nonrwa.2008.11.018. URL http://www.sciencedirect.com/science/article/pii/S146812180800299X.

H. L. Wei and S. A. Billings. Identification of time-varying systems using multiresolution wavelet models. *International Journal of Systems Science*, 33(15):1217–1228, January 2003. ISSN 0020-7721, 1464-5319. doi: 10.1080/0020772031000081982. URL http://eprints.whiterose.ac.uk/1973/.

H. L. Wei, J. Liu, and S. A. Billings. Time-varying parametric modelling and time-dependent spectral characterisation with applications to eeg signals using multiwavelets, April 2008. URL http://eprints.whiterose.ac.uk/74634/.

X. Wen, X. Xu, and Y. Cai. Least-squares wavelet kernel method for regression estimation. In L. Wang, K. Chen, and Y. S. Ong, editors, *Advances in Natural Computation*, number 3610 in Lecture Notes in Computer Science, pages 582–591. Springer Berlin Heidelberg, January 2005a. ISBN 978-3-540-28323-2, 978-3-540-31853-8. URL http://link.springer.com/chapter/10.1007/11539087_74.

Xiangjun Wen, Yunze Cai, and Xiaoming Xu. Wavelet support vector machines and its application for nonlinear system identification. In J. Wang, X. F. Liao, and Z. Yi, editors, *Advances in Neural Networks – ISNN 2005*, number 3497 in Lecture Notes in Computer Science, pages 442–447. Springer Berlin Heidelberg, January 2005b. ISBN 978-3-540-25913-8, 978-3-540-32067-8. URL http://link.springer.com/chapter/10.1007/11427445_72.

X. Xu, Z. Y. Shi, and Q. You. Identification of linear time-varying systems using a wavelet-based state-space method. *Mechanical Systems and Signal Processing*, 26:91–103, 2012. ISSN 0888-3270. doi: 10.1016/j.ymssp.2011.07.005. URL http://www.sciencedirect.com/science/article/pii/S0888327011002937.

Q. Zhang. Using wavelet network in nonparametric estimation. *IEEE Transactions on Neural Networks*, 8(2):227–236, 1997. ISSN 1045-9227. doi: 10.1109/72.557660.

Q. Zhang and A. Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6):889–898, 1992. ISSN 1045-9227. doi: 10.1109/72.165591.

Y. Zhu. *Multivariable System Identification For Process Control*. Elsevier, October 2001. ISBN 0-08-043985-3.