Chapter 2
System Identification Techniques

An intellect knowing at any given instant of time, all forces acting in nature, as well as the momentary positions of all things of which the universe consist, would be able to comprehend the motions of the largest bodies of the world and those of the smallest atoms in one single formula, provided it were sufficiently powerful to subject all the data to analysis.

Pierre Simon Laplace (1749–1827)

2.1 Introduction to System Identification

MODELING is the abstraction of a real process to characterize its behavior. Scientific modeling aims to enhance the investigation of phenomena in order to reveal and better understand cause-effect relationships [1]. The model definition given by Eykhoff [2] introduces the concept of "essential aspects": "*... [model] is a simplified representation of the essential aspects of an existing system (or a system to be constructed), which presents the knowledge of the system in a usable form*".

The set of processes in a system determines the behavior of the system. Every process is determined by its physical and chemical properties, which are not always easily known. A model tries to emulate the 'essential aspects' of the system behavior, simplified by choosing the most significant properties. So, **modeling techniques** can be classified as:

- *a priori* modeling, *white-box* or morphological modeling, by making simple experiments to inquire into the physical or chemical laws involved.
- *a posteriori* modeling or *black-box* modeling, by building a model based only on data (data-driven) without having previous knowledge of the system. The model describes how the outputs depend on the inputs, not how the system actually is, and characterizes the system dynamics (delays, speed, oscillations, and others), though the physical interpretation of the results is not straightforward.

D. Chinarro, *System Engineering Applied to Fuenmayor Karst Aquifer (San Julián de Banzo, Huesca) and Collins Glacier (King George Island, Antarctica)*, Springer Theses, DOI 10.1007/978-3-319-08858-7_2

- *grey box* modeling is an intermediate technique when peculiarities of internal laws are not entirely known, so it is based on both insight into the system and on experimental data analysis.

System identification tries to estimate a *black* or *grey* model of a dynamic system based on observing input-output from experimental data. Zadeh (1962) defined *system identification* as: "*... the determination on the basis of input and output, of a system (model) within a specified class systems (models), to which the system under test is equivalent (in terms of a criterion)*".

The availability and reliability of the design techniques of system identification did expand the application fields beyond the scope of industrial applications. As a result, system identification models have been applied in other diverse fields, for example, economy, environment, biology, psychology, biomedical research, hydrology, and glaciology. The identification problem requires a set of model structures, a validation criterion and an aim [3]. Criteria and models will be presented over the course of this chapter. Some examples of identification aims could be listed here:

- To design control strategies for a particular system (e.g., in optimizing an electrical microgrid operation).
- To analyze the properties of the system (e.g., quantity rates in a medication reaction).
- To forecast the evolution of the system (e.g., future climate prediction according a IPCC downscaling model)
- To identify hidden factors influencing a system (e.g., sun spots in the karst spring).
- To improve the internal knowledge of the system (e.g., the delay in the aquifer discharge with respect to precipitation events).
- To identify the interaction between coupled systems (e.g., climate and glaciers).

The objective of this chapter
This chapter surveys the main approaches to identification and analysis of systems and their theoretical basis, in order to set the methodological scene for the study of the experimental cases described in later chapters, especially highlighting the procedures that mostly have been implemented in natural systems.

Outline of this chapter
The first Sect. 2.2 of this chapter deals with problems in the acquisition of data from sensors including the sampling chosen (2.2.1) and treatment of outliers (2.2.2) to reach a sufficient time series quality for the subsequent treatment of the information.

Section 2.3 is a review of the classical time series as the first approach to understand the underlying mechanism in the system. Two scenarios are presented: time domain analysis (2.3.1) with autocorrelation structures, and frequency domain analysis (2.3.2) with an explanation of concepts like Fourier transform and frequency spectrum.

Although wavelet techniques are included in the frequency domain issues, a special section is dedicated to wavelet techniques (2.4), because of its prominent application in the practical cases. The section contains a description of mother wavelets, detailed clarification of wavelet transform (2.4.2), the process of estimation of the wavelet power spectrum (2.4.3), and some discussions on using these techniques (2.4.4).

A review of model structures (2.5) is necessary for the identification process to select suitable, identifiable model structures. Some linear time-invariant models are described in Sect. 2.5.1. Nonlinear models (2.5.3) go through Volterra series (2.5.3.1) and Hammerstein-Wiener models (2.5.3.2).

Section 2.6 deals with several identification techniques, starting with the elucidation of the posed problem (2.6.1) and a simple literature overview (2.6.2).

Parametric identification (2.8) relies on a model previously defined by a set of parameters that must be calculated to accomplish a given quality criteria. Linear techniques are given in (2.8.1) with some advice on selection and verification criteria of the models (2.8.2).

Nonparametric identification methods are described in 2.7, starting with methods in the frequency domain (2.7.2), including classical spectral analysis (2.7.2.1), going on to wavelet cross spectrum (2.7.2.2), and finishing with wavelet coherence fundamentals (2.7.2.3). In the time domain Sect. (2.7.1) there are methods for system identification, including cross correlation (2.7.1.1) and impulse response (2.7.1.2).

A final section provides an overview of nonlinear identification techniques (2.9) which serves as a tour of nonlinear parametric identification (2.9.1) remarking on the main issues of Volterra identification (2.9.1.1) and Hammerstein-Wiener identification (2.9.1.2). Additionally, methods pertaining to nonlinear nonparametric identification are mentioned (2.9).

Finally, in the conclusion Sect. 2.10, important assertions related to this chapter are offered.

2.2 Time Series

To capture critical information about the processes to be investigated, field data are achieved through the sensor network. Process variables should be sampled for a duration and sampling frequency enough to obtain those quality time series that the

analysis requires. The sequence of observations on one variable y(t), t ∈ T (T is the discrete times domain), is called *time series*. The observation are usually equally spaced and indexed by integers (t = 1,…,n) where n indicates the number of observations. The main objective of time series analysis is to get mathematical inferences from the sample data obtained from field sensors.

A lot of problems can be solved by time series as: prediction (e.g., future weather precipitation), identification or abnormal peaks (e.g., outbursts in a glacier discharge), trends (e.g., sea global warming in the last century), etc.

Before performing the analysis of the time series, some issues should be addressed on the form of achieving raw data, check the integrity and reliability of preparing procedures; i.e., the sampling method, the outliers detection, and the lost data recovery.

2.2.1 Sampling Period

In many monitoring applications, there are some problem related to sensor devices to sample environment variables, mainly concerning with computing data as the memory size, processing capability and power supply. Sampling is the process by which continuous time signals—such as air temperature or water levels—are turned into discrete time signals. About batteries, if the sampling frequency is set too high, then the energy consumption would be so high that the sensor battery would be depleted too soon. To avoid this problem, the sampling frequency can be reduced, but this is not always possible. The Nyquist-Shannon sampling theorem states that if a function y(t) contains no frequencies higher than ω_N, it is completely determinable by a sampling process of frequency 2ω. So, the sampling frequency ω_S should be bounded according the Eq. 2.1:

$$2\omega = \omega_N < \omega_S < \omega_C \tag{2.1}$$

where ω_N is the Nyquist frequency and ω_C is the critical frequency for the sensor battery duration. On this matter, Alippi et al. [4] presented an adaptive sampling algorithm for effective energy management in wireless sensor networks. Some of these sensors require computing capability to integrate a distributed artificial intelligence which offers a wide range of possibilities for the operation, automation and control of different systems [5].

The Nyquist-Shannon criterion provides some clearly stated bounds, but it does not assure that a system can operate right at the Nyquist rate. So, some practical caveats should be taken in mind in designing a sampling process. The Nyquist criterion assumes: a sampling is regular (low noise); the value obtained for each sample should have an infinite precision (issue never happens in sampling natural system); there can be no component in the original signal higher than those correspond to half the sampling frequency; and the criterion does not take in account if (so prevalent in system identification) the signal is going to be processed (modified) after sampling [6].

2.2.2 Outliers

Before performing the analysis of the time series, a preparation phase is carried out to check the integrity and reliability of measurements. **Outliers** are the discordant and unexpected values that can appear in a time series, which Hawkins [7] defined as: "*the observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism*". Barnett and Lewis [8] named outliers as the data deviated markedly from other members of the sample in which it occurs.

Then, the goal of the outliers identification is the location of suspicious values to removed or recovered by a more suitable value, in order to avoid contamination and distortion of the underlying probability distribution. Really, it is hard to tell whether a value is an outlier or true data. So, outliers not always must be considered an erroneous data to be rejected. Outliers can be values produced by natural dynamics of the processes to study, but there are outliers that can be due to external factors of the system, as sensor device failure, mistake in entering data, malfunctioning of the datalogger, break in the communication, etc. [9].

Statistics can calculate if the probability of the candidate observation is small or is enough far from the rest, to be considered as an outlier. Among the distinct methods to detect and correct outliers, the Rosner criterion is well known [10]. A generalized extreme studentized deviate (ESD) is the fundamentals in Rosner criterion to detect from 1 to k outliers in a data set that follows an approximately normal distribution. This procedure handle the error both under the hypothesis of no outliers and under the alternative hypotheses of 1, 2,…, k−1 outliers. Rosner's method detects and removes multiple outliers in a single step.

Under the hypothesis that an excessive deviation of the expected spectrum of a signal is interpreted as the possible presence of outliers, a new test of outliers has been proposed and applied by Chinarro et al. [11] to raw data of Fuenmayor spring. It is the Wavelet-Rosner test, as an extension of Rosner test in the frequency domain (Sect. 2.4.2.4).

2.3 Classical Time Series Analysis

Time series analysis definition was given by Tukey [12] : "*Time series analysis consists of all the techniques that, when applied to time series data, yield, at least sometimes, either insight or knowledge, and everything that helps us choose or understand these procedures*". Thus, a time series can unveil some concealed information about the system, as periodicity, outliers and trends, using typical statistics estimators.

Univariate analysis methods characterize the structure of an individual time series—simple analysis—while *bivariate methods* study the relationship between two different time series—cross analysis.

Under the concept "classical time series analysis" two approaches are studied. One is the analysis in the *time domain* based on concepts like autocorrelation and autoregression. Another is the analysis developed in the *frequency domain* that deals with spectral analysis.

2.3.1 Time Domain Analysis

The time series y(t) is a discrete signal expressed as $\{y(kT)\}$ and $\{x(kT)\}$ that is a sequence obtained from the output (analogously x(t) for the input) of a system at sampling instants ($t_k = kT; k = 0, 1, 2, \ldots, N$), where T is sampling interval and N the number of samples.

Autocorrelation refers to the correlation of a time series with its own past and future values. In order to determine the expression for autocorrelation, the following functions over a discrete time series $y(kT) = \{y_1, y_2, \ldots, y_N\}$ are defined below.

(a) Variance:

$$\sigma_y^2 = \frac{1}{N}\sum_{k=0}^{N}(y(kT) - \mu)^2 \tag{2.2}$$

where μ is the mean of N elements in the time series $y(kT)$.

(b) Autocovariance with lag = τ:

$$C_y^{\tau} = \frac{1}{N}\sum_{k=0}^{N-\tau}[(y(kT) - \mu)\,(y(k+\tau)T - \mu)] \tag{2.3}$$

(c) Autocorrelation as the rate between the variance and the autocovariance:

$$R_y^N = \frac{C_y^{\tau}}{\sigma_y^2}; \quad \text{for any } \tau : R_y^N(\tau) = \frac{C_y(\tau)}{\sigma_y^2} \tag{2.4}$$

Coefficient R_y^N at lag τ given in Eq. 2.4 defines the autocovariance (Eq. 2.3) normalized by the variance (Eq. 2.2) [13]. Positive autocorrelation might be considered a specific form of 'persistence' of events, and time series is better predictable because future values depend on current and past values.

2.3.2 Frequency Domain Analysis

A useful tool to characterize signals is the power density spectrum, due to spectral analysis is concerned with estimating the relative importance of different frequency

bands in a signal. The spectrum from a finite-length sequence of samples $y_N(kT)$, with (K = 0,1,…,N), is given by Eq. 2.5:

$$\phi_y^N(n) = \frac{1}{N}|Y_N(n)|^2 \tag{2.5}$$

where $Y_N(k)$ is the Discrete Fourier Transform (DFT) of the time series $y_N(kt)$ and calculated by:

$$\mathscr{F}\{y_N(kT)\} = Y_N(n) = \sum_{k=0}^{N-1} y_N(kT)\, e^{-(i2\pi nk/N)}$$

As $Y_N(n)$ is a discrete time signal, the frequency variable is also discrete $\omega = n\frac{2\pi}{N}$. Each $Y_N(n)$ is a complex number that encodes both amplitude and phase of a sinusoidal component of function $y_N(kT)$. The sinusoid frequency is k/N cycles per sample. $\phi_y^N(n)$ is called the *periodogram* of $y_N(kT)$. Periodogram analysis was proposed by Stokes [14] and applied by Schuster [15] to analyze sunspot data.

However, this estimate is very fluctuating and only gives a rough picture of the power density spectrum. See Wellstead [16] or Ljung and Glad [3] for an explanation of this problem. A first solution is to average a number of periodograms calculated over different segments of the full signal sequence. It is called the Welch's method [17]. The power density spectrum is the Fourier transform of the autocorrelation, as expressed in Eq. 2.6. See, for example, Ljung and Glover [18] or Ljung and Glad [3] for the basis of this estimate.

$$\phi_y^N(n) = \sum_{k=-N+1}^{N-1} R_y^N(k) e^{-(i2\pi nk/N)} \tag{2.6}$$

The most extended method to power density spectrum is the Blackman-Tukey procedure [19]. The idea is to smooth the periodogram by averaging over a number of neighboring frequencies using a windowed technique. For this last technique, another estimate of the power density spectrum is normally used:

$$\phi_y^N(n, W) = \sum_{k=-M+1}^{M-1} W(k) R_y^N(k) e^{-(i2\pi nk/N)} \tag{2.7}$$

where $W(k)$ is called the lag window and M is the window size that should be small compared with N. There are several window functions, but the most common used in spectral analysis is the Hamming window [3]. In hydrology, it is frequently necessary to numerically evaluate the real part of the *one-sided* Fourier transform of of autocorrelation. The expression of the spectral density function of a signal is given by:

$$S_f = 2\left[1 + 2\sum_{k=1}^{N} W(k) R_y^N(k) \cos 2\pi f k\right] \tag{2.8}$$

where W(k) is the lag window, commonly a Hamming window in spectral analysis.

2.4 Wavelet Transform Techniques

Wavelet techniques supply solutions to the time-scale analysis by decomposing a signal into a superposition of scaled and shifted versions of an original wavelet (also called mother) with specific properties, as a fast-decaying oscillating function. During the last 15 years, great strides in the development of the theory of wavelets have made. The search continues in new application areas and theoretical approaches. The literature devoted to wavelets is very voluminous, so a strict selection of fundamentals is treated in the next sections.

Wavelet theory emerges in the mid-1980s by Grossman and Morlet [20] who apply Gabor wavelets to model echo signals for underground oil prospecting. Jean Morlet was prospecting oil for the Elf-Aquitaine company, sending pulses to underground and analyzing their echoes by Fourier transforms; giving that, the high frequencies of the echoes correspond to thin layers and the low frequencies to the thicker ones. Nevertheless, he found that a lot of reflected signals corresponding to the different layers, were interfering with each other, and he can not separate the required ones. If he selects an extremely small windows (Windowed Fourier Transformation (WFT) [21]) to analyze only signals in high frequencies, all information about low frequencies would be lost. Morlet had to manage the problem in a different way. Instead of fixing the size of the window, he kept constant number of oscillations in the window and varied the width of the window, by stretching or compressing. This made possible to decompose signals simultaneously by time and frequency, giving useful information: *what frequency and when it is produced*. The wavelet engineering was born [22].

In a **historical overview** (Table 2.1), the wavelet theory starts from methods that have been essential in the development of engineering during almost two centuries. Haar [23] introduced the first compactly supported family of functions. After a long time elapse without contributions, Gabor [21] introduces a family of non-orthogonal wavelets with two components, a complex sinusoidal carrier and a Gaussian envelop. A wavelet filter banks for decomposition and reconstruction of a signal were introduced by Esteban and Galand [24], although Crochiere et al. [25] roughly introduce the same idea in speech acoustics, called sub-band coding.

The literature rapidly spread out and wavelet analysis is now used extensively in physics, geophysics [26], economy [27], epidemiology, neuroscience, signal processing Ricker [28], hydroclimatology [29], oceanography [30], hydrogeology [31], hydrology [32], electricity demand [11, 33], remote sensing data [34], computing complex problem as Maxwell's curl equations [35], and other fields.

Table 2.1 History of wavelets

Year	Authors	Facts
1807	J.B. Fourier	Any periodic function can be expressed as an infinite sum of sine and cosine waves of different frequencies. His ideas faced much criticism from Lagrange, Legendre and Laplace for lack of mathematical rigor and generality, and his papers were denied to be published, until 15 years later
1909	A. Haar	He discovers a "base" of functions that are now recognized as the first wavelets. They consist of a short positive pulse followed by a short negative pulse
1930	J. Littlewood and R. Paley	Local information on a wave, as the duration of a pulse of energy can be recovered by grouping the terms of its Fourier series in "octaves"
1946	D. Gabor	He adapted the Fourier Transform to analyze only a small section of the signal at a time. Gabor's adaptation, called Short-Time Fourier Transform (STFT), decomposes a signal into a two-dimensional function of time and frequency. The result is a "packets time-frequency" or "frequency Gabor"
1960	A. Calderon	He provides a mathematical formula that allows mathematicians subsequently recover a signal from its wavelet expansion
1976	D. Esteban and C. Galand	They recognize the subband coding, a way of encoding digital transmissions for telephone
1981	J. Morlet	He discovers a way to decompose the seismic signals where called *wavelets of constant shape*. Ask for help to Alex Grossmann, quantum physicist, to show that the method works
1982	E. Adelson and P. Burt	They developed the "pyramidal algorithm" for image compression
1984	Morlet and Grossmann	They first introduced the term "wavelet" in mathematical language
1985	Y. Meyer	Before 1985, a lot of researchers thought that there was no orthogonal wavelet except Haar wavelet. Yves Meyer constructed the second orthogonal wavelet called Meyer wavelet (soft orthogonal wavelets)
1986	S. Mallat	Shows that the Haar basis, the Littlewood-Paley octaves, frequencies and filters Gabor, and subband from Galand and Esteban, are all related to algorithms based on wavelets
1987	I. Daubechies	She constructs the first smooth orthogonal wavelets with a solid foundation and systematical method, to be used as a practical tool that any scientist can easily program and operate
1988	S. Mallat and Y. Meyer	Mallat and Meyer proposed the concept of multiresolution
1990	D. Donoho and I. Johnstone	They used wavelets to "remove noise" of images, making them even sharper than the originals
1992	T. Hopper, J. Bradley and C. Brislawn	They developed a method based on wavelet to compress its huge database of fingerprints that were applied in Criminal Information Services of FBI

(continued)

Table 2.1 continued

Year	Authors	Facts
1995	Pixar Studios	Presentation of film Toy Story 2, where some forms are provided by subdivision surfaces, a technique mathematically related to wavelets
1996	W. Sweldens	He introduced a new technique, the so-called lifting scheme, which became the basic tool of second generation wavelets
1999	The International Standards Organization	Published a new digital image compression called JPEG-2000. The new standard uses wavelet to compress image files in a ratio of 1:200, without appreciable loss in image quality

Source Daubechies et al. [36] and information gathered from wavelet literature

2.4.1 Wavelet Function

The wavelet transformation needs a basis function—like Fourier transform requires sinusoid basis function, though in wavelets, there are a wide range of basis function families. Principles and conditions of this basic functions are related below.

$L^2(\mathbb{R}^d)$ is a Hilbert space and denotes the set of square integrable functions, i.e., the set of functions defined on the real line such that $\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty$. Since this integral is usually referred to as the energy of the function x, this space is also known as the *space of functions with finite energy*. Therefore, $L^2(\mathbb{R}^d)$ has an inner product $\langle x, y\rangle = \int_{-\infty}^{\infty} x^*(t) \cdot y(t)dt$, and an associated norm $\|x\| = \langle x, y\rangle^{1/2}$.

The term wavelets refers to a set of small waves formed by dilations and translation of a single function $\psi(t)$ which should be square integrable over the range of real time of space $L^2(\mathbb{R})$. Table 2.2 summarizes the requirements to be met by a wavelet [37].

The function $\psi(t)$ is called "mother wavelet" or "basic wavelet" while the dilated and translated functions derived from the "mother wavelet" are called "daughter wavelets" or simply "wavelets" (Fig. 2.1). These daughter wavelets have the same shape as their mother wavelet. Their amplitude should rapidly decay away from the center of the wave in both time and frequent domains. The functional relationship between daughter $\psi_{s,\tau}(t)$ and mother $\psi(t)$, in the scale s and displacement τ, is

Table 2.2 Requirements of a wavelet function $\psi \in L^2(\mathbb{R}^d)$ ($\hat{\psi}$ is the Fourier transform of the wavelet function ψ)

	Description	Condition
a	The average value of the wavelet in the time domain should be zero	$\int_{-\infty}^{\infty} \psi(t)dt = 0$
b	The function must have finite energy	$\int_{-\infty}^{\infty} \|\psi(t)\|^2 dt = 1$
c	Admissibility . The inverse wavelet transform only exists for $0 < C_\psi < \infty$. This means that the analyzed signal can be reconstructed without loss of information. The constant C_ψ is called the admissibility constant	$C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{\left\|\hat{\psi}(\omega)\right\|^2}{\|\omega\|} d\omega$

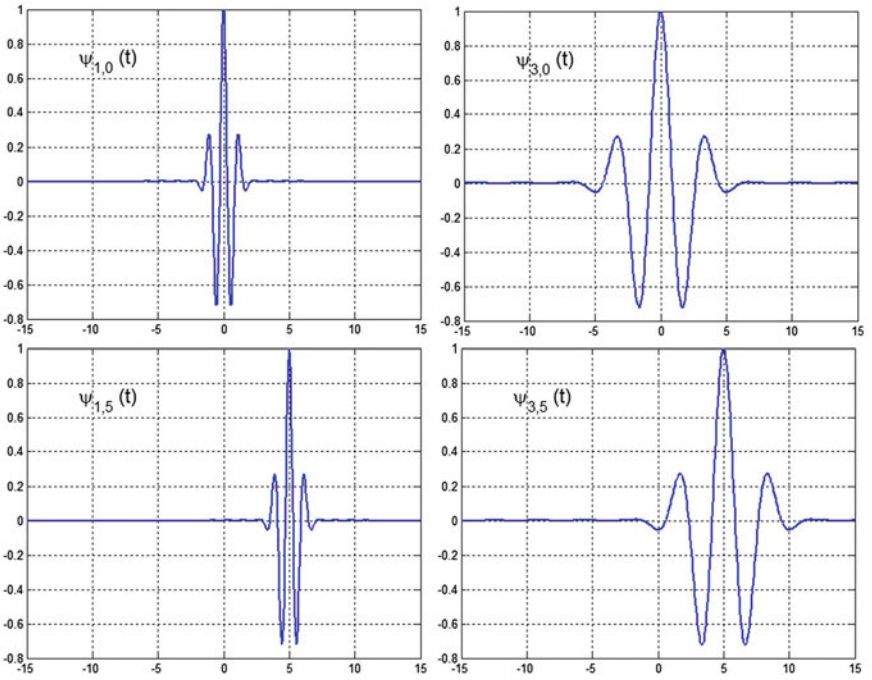


Fig. 2.1 Morlet's "wavelet daughters" $\psi_{s,\tau}(t)$, that are dilated and translated functions derived from the "mother wavelet" $\psi(t)$, according with Eq. 2.9

expressed as (Eq. 2.9):

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(t)\left(\frac{t-\tau}{s}\right) \tag{2.9}$$

where s, τ are real and s>0. Wavelets expressed by (2.9) include an energy normalizatation $s^{-1/2}$ which keeps the energy of the daughter wavelets the same as the energy of their mother.

Morlet wavelet is the most popular complex wavelet used in practice, which mother wavelet is defined as Eq. 2.10 and represented in Fig. 2.1

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}}\left(e^{j\omega t} - e^{-\frac{\omega^2}{2}}\right)e^{-\frac{t^2}{2}} \tag{2.10}$$

where ω is the central frequency of the mother wavelet. Note that the term $e^{-\frac{\omega^2}{2}}$ is used for correcting the non-zero mean of the complex sinusoid, and it can be negligible for $\omega < 5$

For practical purposes, for large ω, e.g., $\omega > 5$, Eq. 2.10 can be simplified [38] by taking a complex cos wave modulated by a Gaussian envelope (Eq. 2.11):

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}} e^{j\omega t} e^{-\frac{t^2}{2}} \tag{2.11}$$

Morlet wavelet provides a better energy localizing and higher frequency resolution, although the frequency-coordinate window shifts along frequency axis as scaling.

2.4.2 Wavelet Transforms

The basic aim of wavelet analysis is both to determine the frequency (or scale) content of a signal and to assess and determine the temporal variation of this frequency content [39]. Although wavelet analysis covers a wide range of methods and applications, fundamental operations are *wavelet transforms*, which are appropriate for the many natural phenomena that have the property that high frequency events happen for short durations.

2.4.2.1 Continuous Wavelet Transform (CWT)

A wavelet transform correlates the signal with a family of waveforms $\psi_{s,\tau}$ or *wavelets*—it is also called time-frequency atoms by Mallat [40] and kernel by other authors—that meet the conditions of Table 2.2. The corresponding continuous wavelet time-frequency transform of $f \in L^2(\mathbb{R})$ is expressed by (Eq. 2.12):

$$\mathscr{W}_{\psi,s,\tau}\{f(t)\} = \int_{-\infty}^{\infty} f(t) \cdot \psi^*_{s,\tau}(t)dt = \langle f, \psi_{s,\tau} \rangle \tag{2.12}$$

where $\mathscr{W}_{\psi,s,\tau}\{f(t)\}$ is the CWT of f(t) with basis function family $\psi_{s,\tau}$.

The wavelet coefficients represent a measure of similarity in the frequency content between a signal and a chosen wavelet function. These coefficients are computed as a convolution of the signal and the scaled wavelet function, which can be interpreted as a dilated band-pass filter because of its band-pass like spectrum.

2.4.2.2 Continuous Wavelet Transform with Discrete Coefficients (CWTDC)

Signals are usually band-limited, which is equivalent to having finite energy, and therefore just a constrained interval of scales is useful. However, the continuous wavelet transform produces redundant information when capturing all the characteristics of the signal. The Discrete Wavelet Transform (CWTDC) has been created to minimize the redundancies produced by CWT.

It is possible to compute the wavelet transform for just a proper selection of values of the frequency and time parameters and still not loose any information as recovering the original time series from its transform. The continuous wavelet transform with

discrete coefficients is very similar to the continuous wavelet transform, but the parameters s and τ are fixed to the power of 2, following these expressions:

$$\tau = 2^{-\mathrm{j}}\mathrm{n},\ s = 2^{-\mathrm{j}}, \text{for m} \geq 0 \text{ and n} \in (-\infty, \infty) \tag{2.13}$$

The functional relationship between daughter $\psi_{k,j}(t)$ and mother $\psi(t)$, in the scale k and displacement j, is expressed as:

$$\psi_{k,j}(t) = 2^{j/2}\psi(2^j - k) \tag{2.14}$$

So, Eq. 2.14 is a special case of Eq. 2.9. From the continuous wavelet function (Eq. 2.12) and the new values of s and τ from Eq. 2.13, CWTDC takes the form of Eq. 2.15.

$$\mathscr{W}_{\psi,k,j}\{f(t)\} = \sqrt{2^j} \int_{-\infty}^{\infty} f(t) \cdot \psi^*_{k,j}(2^j t - k)dt \tag{2.15}$$

2.4.2.3 Discrete Wavelet Transform (DWT)

Although the output of the continuous wavelet transform contains discrete coefficients, its implementation could be hard, since the input signal is continuous. Discrete wavelet transform is the alternative.

If f $\in L_2(\mathbb{R})$ and h $\in L_1(\mathbb{R})$, the convolution between the two signals: $g(x) = (f \otimes h)(x) = \int_{-\infty}^{\infty} f(t)h(x-t)dt$. The continuous wavelet transform of a f(t) signal, given at Eq. 2.12, yields a infinite set of wavelet coefficients. In a discrete form, where a time series u is a discrete sequence values of $(u_1, u_2, \ldots, u_n)$ separated in time by a constant time interval δt, the expression for the wavelet coefficient is given at time index j and scale s in the expression 2.16.

$$\mathscr{W}_{\psi,s,j}(u_n) = \sum_{n=0}^{N-1} u_n \psi^* \left[\frac{(n-j)\delta t}{s}\right] \tag{2.16}$$

where u_n is the discrete sequence, N denotes the length of the studied time series, ψ^* is the wavelet complex conjugated, and δt denotes the sampling period.

The algorithm to calculate the wavelet transform from Eq. 2.16, is a loop of convolutions performed N times for each scale, where N is the number of elements in the time series.

The convolution theorem states that the Fourier transform of the convolution of two functions is the product of Fourier transforms of each function $\mathscr{F}\{u(t) \otimes y(t)\} = \mathscr{F}\{u(t)\} \otimes \mathscr{F}\{y(t)\}$. Performing the Fourier transform in both sides of the Eq. 2.16, the inverse of wavelet transform can be expressed as Eq. 2.17. Applying the Fourier inverse transform to $\mathscr{F}\left\{\mathscr{W}_{\psi,s,j}(u_n)\right\}$, and following the work of Grinsted et al. [26],

an efficient algorithm has been created to compute the discrete wavelet transform for the natural time series of this dissertation. This DWT can be computed by the inverse Fourier transform, as the Eq. 2.17 states.

$$\mathcal{F}\left\{\mathcal{W}_{\psi,s,j}(u_n)\right\} = \sum_{k=0}^{N-1} \hat{u}_k \cdot \hat{\psi}^*(s\omega_k) e^{i\omega_k n\delta t} \tag{2.17}$$

where k = 0 …N−1 is the frequency index, the Fourier transform of the function $\psi(s,t)$ is $\hat{\psi}(s\omega_k)$. The angular frequency ω_k admits the following values:

$$\omega_k = (2\pi k)/(N\delta t) \text{ for } k \leq N/2$$
$$\omega_k = -(2\pi k)/(N\delta t) \text{ for } k > N/2$$

2.4.2.4 Ouliers Detection in the Frequency Domain

Since outlier is an observation with characteristics of high-frequency phenomena, then wavelet technique is an excellent tool as outlier detector because of good location of frequencies. Most of energy and information of the data are usually concentrated in the first few coefficients. The outliers, as the noise, are on high frequency bands and reside in high-order coefficients. Therefore, the true data and outliers can be separated in the wavelet space (frequency domain).

In a discrete wavelet analysis, a signal u(t) can be represented by a decomposition of the signal into approximations A_j and detailed coefficients D_j. This is accomplished using shifted and scaled versions of the original (mother) wavelet as given in Eq. 2.16. In practice, the wavelet coefficients are computed efficiently using the *pyramid algorithm*, introduced in the context of multiresolution analysis by Mallat [40], that is based on a pair of high and low pass filters. The DWT in Eq. 2.16 produces a matrix of coefficients $[c_{i,k}]$. Then, the function can be represented by:

$$u(t) = \sum_{i=-\infty}^{j} \sum_{k} c_{i,k}\psi_{i,k}(t) = \sum_{i=-\infty}^{j-1} \sum_{k} D_{i,k}\psi_{i,k}(t) + \sum_{j} A_{j,k}\psi_{j,k}(t) \tag{2.18}$$

A_i is the approximation component set of the time series, and represents the low-frequency content. This low pass filter is like to continuously calculate a moving average of weighted data. D_i is the detailed component set of the time series, and represents the high-frequency content. This high pass filter consists on a moving difference of the data. These two sets of wavelet coefficients facilitate the recursive form of the *pyramid algorithm* [40].

Struzik and Siebes [41] propose a methodology capable of determining the statistical nature of the non-stationary process. The method checks the internal consistency of the scaling behavior of the process within the paradigm of the multifractal spectrum. Deviation from the expected spectrum is interpreted as the potential presence

of outliers. Chinarro et al. [11] proposed a *wavelet-rosner test* and applied it to the time series of a karst aquifer. The method also stems from the wavelet based multiresolution analysis. After checking the A_i and D_i coefficients (Eq. 2.18) to detect outliers, Rosner's test is applied to remove or replace the abnormal values. A brief description of method in steps is:

1. Firstly, to avoid null values, a temporary shrunk series has been created with all elements from the raw data except null elements.
2. Transform the time series to the wavelet domain. DWT decomposes the time series and computes the approximation coefficients vector A_i and detailed coefficients vector D_i, at level 1, following (2.18).
3. Apply the Rosner test on the A_i coefficients to get outliers in the frequency domain. On this step, another method to remove outliers can be used, but Rosner's test has been weel tested in hydrological series.
4. Eliminates all outliers from the A_i and from analogous index in D_i, then two shrunk vector A_r and D_r are created.
5. To restore the time series, use A_r and D_r to compute the inverse wavelet transform.

2.4.3 Wavelet Power Spectrum (WPS)

The wavelet power spectrum helps to estimate the repartition of energy in the signal to determine the concentration of a signal in singular instants and frequencies. Temporal variation in the distribution of energy across scales is one of the most usual applications of the wavelet transform.

The Wiener-Khinchin theorem states that energy spectral density of a function is the Fourier transform of the corresponding autocorrelation sequence [28]. Analogously, wavelet power spectrum ($\mathscr{P}_{\psi,s,\tau}$) of signal u(t) is defined as autocorrelation function of the wavelet transformation ($\mathscr{W}_{\psi,s,\tau}$) of u(t), and describes the power of the signal u(t) at a certain time τ on a scale s:

$$\mathscr{P}_{\psi,s,\tau}\{u(t)\} = \mathscr{W}_{\psi,s,\tau}\{u(t)\} * \mathscr{W}^{*}_{\psi,s,\tau}\{u(t)\} = \left|\mathscr{W}_{\psi,s,\tau}\{u(t)\}\right|^2 \tag{2.19}$$

Because the wavelet function $\psi(\tau)$ is in general complex, the wavelet transform $\mathscr{W}_{\psi,s,\tau}$ is also complex, with a real part, $\Re(\mathscr{W}_{\psi,s,\tau})$, an imaginary part, $\Im(\mathscr{W}_{\psi,s,\tau})$, an amplitude $\left|\mathscr{W}_{\psi,s,\tau}\right|$, and a phase, $\Re(\mathscr{W}_{\psi,s,\tau})/\Im(\mathscr{W}_{\psi,s,\tau})$. Analogously, WPS can be expressed in the same components.

Wavelets add a new dimension in the spectral analysis to work simultaneously with time and frequency. Wavelet power spectrum is, in fact, a three-dimensional depiction, with time on the x-axis, frequency or scale on the y-axis, and the z-axis is to render the power magnitude at a particular time and frequency. This is a suitable tool for the spectral analysis of a non-linear system.

2.4.4 Wavelet Transformation Caveats

2.4.4.1 Cone of Influence (COI)

When applying the CWT to a finite length time series, the scalogram inevitably suffers from border distortions. The cause is that the values of the transform at the ends of the series cannot be accurately calculated because the transform calculus takes values outside the series range. These hedge effects also increase with s in a rate that depends on the mother function. The region in which the transform suffers from these edge effects is called the cone of influence (COI) and should be marked to take care in interpreting the belonged values [42]. One solution is to pad the end of the time series with zeroes before applying the wavelet transform and then remove them afterward. The padding is an extension of time series should be sufficient to spread out the time series to the next power of two. The zero padding reduces the variance, but introduces discontinuities at the endpoints and decreases the amplitude near the edges as going to larger scales [43].

As wavelet coefficients at COI suffer the same input discontinuity, the solution may be to rescale the remaining wavelet with choosing the *e-folding* time, i.e., as the distance at which the wavelet power drops by a factor e^{-2}. Larger e-folding time implies more expansion of the wavelet spectrum, The e-folding time is a measure of the wavelet width, relative to the wavelet scale s and ensures that the edge effects are negligible above a threshold for a given signal u(t) [26].

2.4.4.2 Choosing Wavelet Function

One singular characteristic of wavelet analysis is the arbitrary choice of the wavelet function. The below list is based on factors given by Torrence and Compo [43], in order to select a suitable wavelet to get best performance, and has been completed with other considerations by the author of the thesis:

(a) *Discrete or continuous.* DWT provides a more compact representation of data. So, DWT is rather suited for image processing, signal coding, noise reduction and computer vision. Nevertheless, CWT (also CWTDC) is a transformation that provides a high redundancy of data, and is suitable for time series analysis and feature extraction purposes.

(b) *Orthogonal or nonorthogonal.* The use of an orthogonal basis implies the use of DWT while a non-orthogonal wavelet function can be used with either the discrete or the continuous wavelet transform. Orthogonal wavelet functions have a zero correlation each other while non-orthogonal wavelets have a nonzero correlation. Using an orthogonal wavelet, the signal can be transformed to the frequency domain and then return to the time domain with a negligible loss of information. Orthogonal wavelet analysis is useful for signal processing because it gives the most compact representation of the signal. Non-orthogonal wavelets

tend to surplus energy, because of overlapping, and require a normalization to optimize the acquisition of information. They are useful for time series analysis.

(c) *Complex or real.* A complex wavelet function will return information about both amplitude and phase and is better adapted for capturing oscillatory behavior. A real wavelet function returns only power, but is useful in location of peak frequency.

(d) *Width*. A wide wavelet function will give good frequency resolution and a loss of time resolution while a narrow wavelet function will yield good time resolution but a poor frequency resolutions.

(e) *Shape.* The wavelet function should reflect the type of features to be presented. For time series with sharp jumps or steps, a box-like function would be better such as Harr's wavelet proposed by Haar [23]. Nevertheless, for smoothly varying time series, the recommendable wavelet is a smooth function such as a damped cosine. If a wavelet power spectra has to be performed, then "the choice of wavelet function is not critical, and any function will give the same qualitative results as another" [43].

(f) *Choice of scales.* In orthogonal wavelet analysis, the set of scales s is limited [44]. In non-orthogonal wavelet analysis, an arbitrary set of scales can be built up for a more complete plotting.

To analyze natural systems, this dissertation has chosen the Morlet wavelet in most cases, because of five interesting properties:

- The peak frequency, the energy frequency and the central instantaneous frequency of the Morlet wavelet are all equal facilitating the conversion from scales to frequencies.
- Heisenberg Box area has a reduced size with this wavelet, i.e., the uncertainty reaches a minimum value. Then Morlet wavelet has an optimal joint time-frequency concentration.
- The time radius and the frequency radius are equal; therefore, this wavelet represents the best compromise between time and frequency concentration.
- Finally, Morlet is a wavelet transformation that yields complex coefficients, with information on both the amplitude and phase. This facilitates the study of gaps and delays between two time series.

2.4.4.3 Advantages of Wavelet Transform

Some features can be observed in the application of wavelet transform and according to Strang [45] and Perrier et al. [46].

Nonlinearity. Analysis with Fourier transform is not completely successful in all types of problems. Exceptions are nonlinear systems, with very brief signals or sudden changes, as the typical time series given in a karst system and glacier discharge; hence, the study of their behavior should be carried out with different tools.

Stationarity. Most traditional mathematical methods that examine periodicities in the frequency domain, such as Fourier analysis, have implicitly assumed that the underlying processes are stationary. The wavelet transform is suitable for the analysis of non-stationary signals because it provides a better time and frequency localization properties, expanding time series into time frequency space, such that the intermittent periodicities can better be localized [46].

Global properties. A Fourier transform hides information about time. It proclaims unequivocally how much of each frequency a signal contains, but is unknowable about when these frequencies were emitted. Therefore, in Fourier transform, any instant of a signal is similar to any other, even if the signal is as complex as public clap echoes in a theater, or changes as radically as the runoff through a river after a severe rainstorm. For an application, wavelets only capture the local time-dependent properties of data; whereas Fourier transforms, due to space-filling nature of the trigonometric functions, can only capture global properties [47].

Computational efficiency Using the *big* O notation [48], the computational complexity of the discrete Fourier transform is $O(n^2)$, where n is a number of time samples. Fourier transform is overtaken by the Fast Fourier transform (FFT) with a complexity O(nlogn) which takes less steps to solve an instance of the same problem. However, this is still under the complexity function O(kn) for discrete wavelet transform in decomposition and reconstruction processes (Fig. 2.2). The transform upshot with wavelets can be implemented in a computer by a quicker and more efficient algorithm.

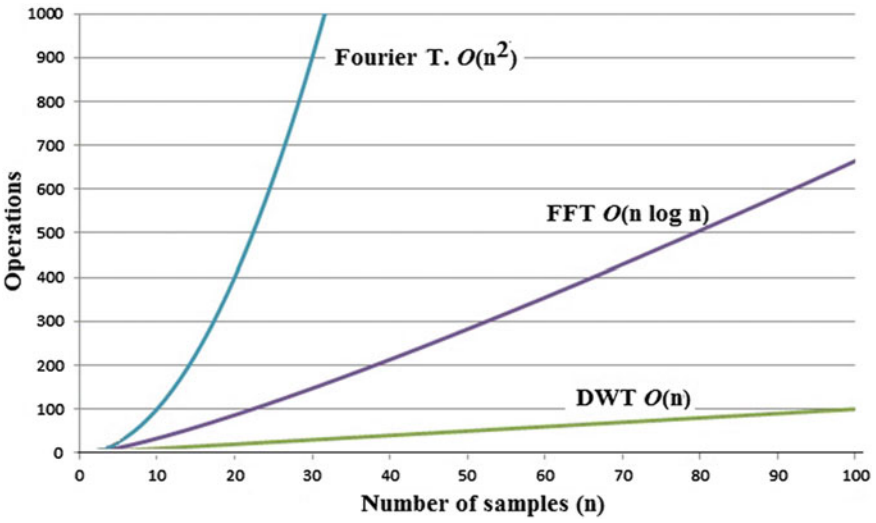


Fig. 2.2 Big-O complexity for Fourier Transform, FFT and DWT

2.5 Models

2.5.1 Linear Time-invariant (LTI) models

Most of the properties of linear time-invariant (LTI) systems are due to the fact that the system can be represented by linear differential (or difference) equations. Such properties include impulse response, convolution, duality, stability, scaling, etc. The properties of linear, time-invariant system should not in general apply to nonlinear systems. Nevertheless, LTI could be a first approach to identify the non-linear system.

The effect of any invariant linear system (LTI) on an arbitrary input signal is obtained by convolution of the input signal with the system's impulse response function. In a LTI system, the output of the system y(t) for an input x(t) can be obtained by the convolution integral:

$$y(t) = g(t) \star x(t) = \int_0^t g(t-\tau)x(\tau)d\tau \tag{2.20}$$

where g(t) is the *impulse respons*e of the system. That is, g(t) is the output of the system with an input x(t) = $\delta(t)$, where $\delta(t)$ is the Dirac delta. The impulse response completely characterizes the dynamic behavior of the system.

Applying the Laplace transform to the convolution integral (Eq. 2.20) we obtain Eq. 2.22 :

$$\mathscr{L}[y(t)] = \mathscr{L}[g(t) * x(t)] = \mathscr{L}[g(t)]\mathscr{L}[x(t)] \tag{2.21}$$

or in simple expression:

$$Y(s) = G(s)X(s) \tag{2.22}$$

where Y(s), G(s) and X(s) are the Laplace transforms of y(t), g(t) and x(t) respectively.

A **Transfer Function** (TF) is the mathematical representation of the relation between the input and output of a system. In a LTI system, TF can be expressed as the ratio of the Laplace transform of the output and the input, and corresponds to the Laplace transform of the impulse response G(s).

$$G(s) = \frac{Y(s)}{X(s)} \tag{2.23}$$

The transfer function of a system is rational fraction with numerator and denominator polynomials of the complex variable s:

$$G(s) = \frac{b_m s^m + \cdots + b_0}{a_n s^n + \cdots + a_0} = \frac{N(s)}{D(s)} \tag{2.24}$$

The roots of N(s) are called *zeros* of the system and roots of D(s) are called *poles* of the system. Poles and zeros are complex numbers that determine the dynamic behavior of the system. The real part of the poles defines the temporal prevalence of the term and the imaginary part its oscillatory behavior. The Transfer Function also characterizes the frequential behavior of a system, that is, how it responds to signals with different frequency components.

Most technological or natural systems are continuous and the signals are involved in dense time. However, when signals have to be processed by a computer in order to monitor, communicate or control a given system, they must be sampled at discrete points in time. The time interval between two sampling instants is the sampling interval or period (T). The models describing the behavior become discrete in time, and the discrete transfer function can be expressed by the *Z transform*.

$$\mathscr{Z}\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \tag{2.25}$$

The concept of **Z transform** plays the same role for discrete time or sampled systems as the Laplace transform does for continuous time systems. Let $x^*(t) = \{x(kT)\}$ and $y^*(t) = \{y(kT)\}$ be the sequences obtained from the input and the output of the system at sampling instants ($t_k = kT; k = 0, 1, 2, \ldots$). The relationship between both data sequences is the convolution sum.

$$y^*(t) \sum_{k=0}^{\infty} x(kT) g_T^*(t - kT) = \sum_{k=0}^{\infty} g_T(kT) x^*(t - kT) = x^*(t) \star g_T^*(t) \tag{2.26}$$

where $g^*(t) = \{g_T(kT)\}$ is called the discrete impulse response of the system with sampling interval T. The sequence $g^*(t)$ is related with the impulse response of the continuous system g(t), but it is not the result of their sampling.

Linear systems are also modelled by discrete LTI Transfer Function, defined as the ratio of the *Z transform* of the output and the *Z transform* of the input. This function is rational with numerator and denominator polynomials of the complex variable z:

$$G_T(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + \cdots + b_m z^{-m}}{a_0 + \cdots + a_n z^{-n}} \tag{2.27}$$

Thus, the activity of the system can be described as a set of parameters a_i, b_i of Eq. 2.27.

A useful format for the discrete LTI transfer function is to describe it in terms of z^{-1} because this is the unit delay operator. Roots of N(z) are called zeros and roots of D(z) are called poles of the system. As in time-continuous systems, the poles and zeros determine the dynamic behavior of the system.

2.5.2 Frequential Transfer Function Model

Let the input of a LTI system be of the form $x = x_0 cos(\omega t)$, it is well known that the response $y(t)$ of the system in steady state is also sinusoidal $y(t) = y_0 cos(\omega t + \phi)$. If $G(s)$ is the transfer function of the system then:

$$y_0 = x_0 |G(j\omega)|$$
$$\phi = arg[G(j\omega)]$$

$G(j\omega)$, that is the transfer function evaluated in $j\omega$, is called the frequential transfer function of the system and the following holds:

$$G(j\omega) = \frac{Y(\omega)}{X(\omega)} \tag{2.28}$$

where $X(\omega)$ and $Y(\omega)$ are the Fourier transforms of input and output, respectively. The frequential transfer function characterizes the frequential behavior of the system, that is, how the frequential components of the input are modified (amplitude change and phase delay) to compose the output. There are some graphical representations of $G(j\omega)$ such as the Bode and Nyquist diagrams that allow a straightforward analysis of the system in function of frequency. Another expression of frequential components of a signal, similar to the Fourier transform, is the power density spectrum, $\phi_x(\omega)$, (signal energy/frequency unit). Equation 2.29 depicts the power density spectrum as the square of the absolute value of its Fourier transform.

$$\phi_x(\omega) = |X(\omega)|^2 \tag{2.29}$$

2.5.3 Nonlinear Models

Nonlinear system identification from input-output data can be performed using general types of nonlinear models such as neuro-fuzzy networks, neural networks, Volterra series or other various orthogonal series to describe nonlinear dynamics.

Non-linear relationships between input and output data provide much flexibility to describe a system. Models to identify nonlinear systems are discussed in Haber [49], and an extensive bibliography classified by nonlinear identification techniques can be found in Giannakis and Serpedin [50]. The use of functional analysis as a tool for the study of nonlinear systems was initially conceived by Wiener [51]. Following this work, Singleton [52] and Bose [53] made pioneering efforts toward engineering applications who left a firm foundation in the theory of functional analysis for both discrete and continuous systems. Other interesting works studied the theory of nonlinear continuous systems using power series Volterra functional [54] and orthogonal expansion of the functional-G Weiner [55]. Many efforts to develop analysis

techniques are related to nonlinear systems, with significant number of people from the Russian school: Liapounoff, Andronov, Chaikin, Kryloff, Bogoliuboff, mentioned by Minorsky [56].

2.5.3.1 Volterra Series

The *Volterra series*, a discrete version of Kolmogorov-Gabor polynomial, was originally developed to describe the nonlinearity of a very general class of nonlinear time-invariant process. Although the Volterra series representation of nonlinearity provides theoretical understanding of nonlinearity, the number of coefficients in this model is excessive and needs enormous requirements on the identification procedure (quality and quantity of data). A large class of nonlinear functionals can be represented in the form of a Volterra series, which maps input signals u to output signals. A nonlinear system can be modeled as an infinite sum of multidimensional convolution integrals of increasing order (2.30).

$$y(t) = \sum_{i=1}^{n} \left| \int_0^{\infty} \right|_i h_i(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} [x(t - \tau_i) d\tau_i] \tag{2.30}$$

where h(τ) is the kernel of convolution. Identifying the nuclei of the Volterra series is the essential problem. In the case n = 1 (a unique term in the development), Volterra series is a first order and corresponds to linear modeling. So, Volterra theory is a generalization of the linear convolution integral approach often applied to linear, time-invariant systems. The behavior of the model depends on the kernels of the integral functionals, and it is these functions that are to be identified. A closely related model was introduced by Wiener, with the form:

$$y(t) = \sum_{n=0}^{\infty} [G_n(k_n, u)](t) \tag{2.31}$$

where the functionals G_n are also integral equations in (2.30), with kernels $k_n = \tau_1, \ldots, \tau_n$ (these were used by Wiener as a sort of orthogonalized version of the Volterra kernels, via a Gram-Schmidt procedure). In discrete-time models, the integrals are replaced by sums.

Nonlinear Volterra theory is widely developed and applied in several fields of science and technology, as an approach to the modeling of nonlinear system behavior. Schetzen [57] goes into more depth about the theory of Volterra and Wiener functionals. Leontaritis and Billings [58] discuss available identification methods for the kernels.

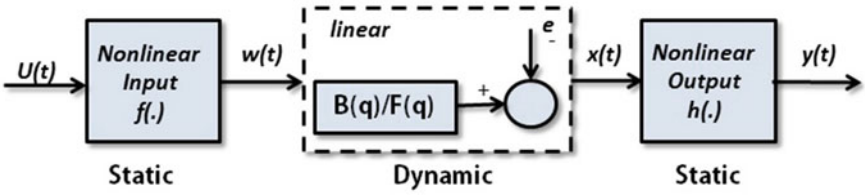


Fig. 2.3 Hammerstein-Wiener model

2.5.3.2 Hammerstein-Wiener Models

An alternative to drive problems of identifying a nonlinear model from input-output data is to use block-oriented nonlinear models consisting of static nonlinear function and linear dynamics subsystem such as Hammerstein model, Wiener model and feedback block-oriented model [59]. When the nonlinear function precedes the linear dynamic subsystem, it is called the Hammerstein model, whereas if it follows the linear dynamic subsystem, it is called the Wiener model.

Block-oriented models can be a flexible alternative in system identification tasks, due to they behold both linear and nonlinear features. Block-oriented models provide structures to study non-linear systems, under the hypothesis of LTI systems with static (no memory) non-linearity.

The model described by Eskinat et al. [60] is an interconnected model in cascade. Wiener model is similar [61], but the order of linear and non-linear blocks are inverted. The feedback block-oriented model consists of a static nonlinearity in the feedback loop of an LTI system [59].

In the Hammerstein-Wiener (HW) model (Fig. 2.3), a linear block model is the central block, represents the system dynamics and can be expressed by an output error polynomial model as Fig. 2.6. The first and last blocks are two nonlinear blocks w(t) = f(u(t)) and y(t) = h(x(t)). Where u(t) and y(t) are the inputs and outputs of the system and w(t) and x(t) are the input and output of the internal linear block. Hammerstein is a submodel with a nonlinear component followed by a linear component. Reversely, Wiener is a submodel with a linear component followed by a nonlinear component (Fig. 2.3). Nonlinear block models represent the static nonlinearities in the system [60] and can be identified following the parametrization given in (2.32).

$$\mathcal{N}(q) = \sum_{i=1}^{n} p_i g_i(x), \quad p = [p_1, \ldots, p_n]^T \tag{2.32}$$

where g_i are a set of specified basis functions such as polynomial expressions, dead-zone, saturation, piecewise, sigmoidnet or wavenet, look-up tables and fuzzy models [62]; and the p_i are the weights.

HW model provides a flexible parameterization for nonlinear models. E.g., a linear model can be estimated, and its quality can be improved, by adding an input or output nonlinearity to the model. On the other hand, HW is easier to implement than others such as neural networks and Volterra models.

2.6 System Identification Problem

The prior sections deal with preparation data (Sect. 2.2), time series analysis (Sect. 2.3), and review of model structures (Sect. 2.5), that are necessary in the identification process.

2.6.1 Posed Issue

The general problem of the system identification is to determine the most suitable model for the system, such as the problem graphically posed in the Fig. 2.4.

The model $g(u(t), \vartheta)$ maps the input u(t) to the output y(t) which is corrupted by perturbations $\varepsilon(t)$. The verification of the model consists of finding the vector ϑ that minimizes the error between real data and predicted data: e(t)= y(t)-$\hat{y}(t)$.

The searching of a model of a system from observed input-output data, implies the **input-output data** in the preparation stage, a set of candidate **model structures**, and some **criteria** to select a particular model in the set (Fig. 2.5).

System identification is carried out through stages of preparation, analysis and the identification as such (selection and optimization). The analysis procedures try to obtain the most possible details inside the system and take useful information from the time series. E.g., the elucidation if it is a linear system or a nonlinear system, a time-invariant system or a time-variant system a continuous system or a discontinuous system, a single input system or a multi input-multi output system, an open-loop system or a closed-loop system, etc (Fig. 2.5).

The selection stage is the identification of a suitable, identifiable model structure. The final identification stage—with the most critical procedures of

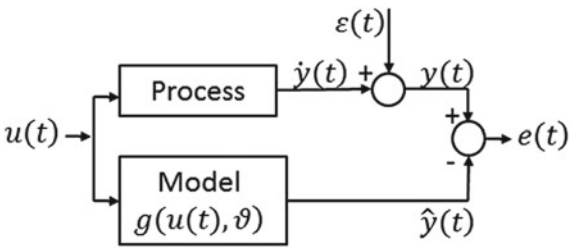


Fig. 2.4 Identification problem

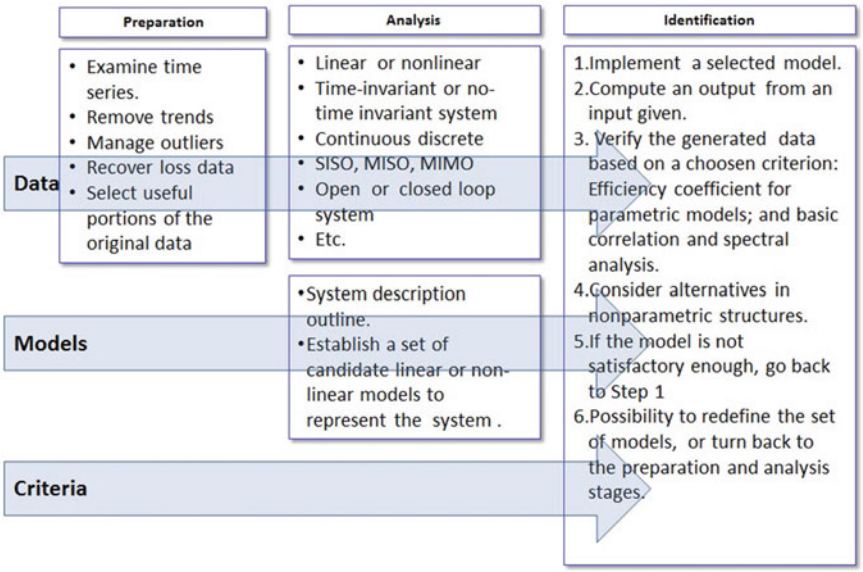


Fig. 2.5 Stages and components in the system identification problem

estimation—computes the best model according to the data and a given criterion. The validation of model measures the ability of the model to explain the observed data different to those used in the model identification and estimation stages, as the general prediction error (maximum likelihood) and efficiency criterion for parametric models; and basic correlation and spectral analysis in nonparametric structures [62]. As each system requires a different model to be chosen among those explained in Sect. 2.5, the process has to be necessary iterative and sometimes applying an *ad hoc* method.

2.6.2 The Literature Highlights

This section can not describe all historical processes on the system identification, only some highlights are mentioned from the engineering views. A further purpose can be found in the work of Deistler [63], with an excellent review of the history of system identification and time series analysis.

Spectrum analysis of time series might have commenced in 1664, when Isaac Newton decomposed a light signal into frequency components as passing the light through a glass prism. In 1800, Herschel measured the average energy in various frequency bands of the spectrum by placing thermometers along each band.

The first foundations on identification processes were set by mathematicians—Gauss (1809) and Fisher (1912)—although with subsequent important contributions

from engineering and statistics. Aström and Bohlin [64] introduced the Maximum Likelihood framework, based on projection techniques in Euclidean space, which have been used extensively on the estimation of the parameters of difference—also known ARMA (AutoRegressive Moving Average) or ARMAX model (AutoRegressive Moving Average with eXogeneous inputs).

Among contributions of Ljung and Glover [18], there is one to clearly separate two independent concepts: the choice of a parametric model structure and the choice of an identification criterion.

In system identification, there are two approaches: parametric and nonparametric identification. In the parametric identification problem, a mathematical structure is assumed to govern the system, and the identification processes is focused only on the determination of unknown parameters for this structure that optimize the representation of the system. Nonetheless, in a non-parametric identification the structure of these equations is also unknown. Nonparametric regression and spectral techniques correspond to this kind of techniques [13, 62].

2.7 Non-parametric Identification

Nonparametric identification techniques provide a very effective and simple way of finding model structure in data sets without the imposition of a parametric one. Commonly, the initial process to carry out is the nonparametric identification, and then, if it were suitable, the parametric identification should be performed. The next sections review the non-parametric identification methods from time domain and frequency domain perspectives.

2.7.1 Non-parametric Identification in the Time Domain

2.7.1.1 Cross-Correlation

Cross-covariance is a non-parametric identification technique and is related with the impulse response g(t) (Eq. 2.20) of a system and thus with its behavior [62, 65]. Assuming that the signals have zero mean, if $y^\star(t)$ and $x^\star(t)$ are uncorrelated, the correlation between the input and the output is:

$$y^\star(t) = \sum_{k=0}^{\infty} g(kT)x^\star(t - kT) + v^\star(t) \tag{2.33}$$

where $v^\star(t)$ is the noise in the system. The signals involved can be regarded as the realization of stochastic processes. We can define the following coefficients and functions: If $v^\star(t)$ and $x^\star(t)$ are uncorrelated, the (cross) covariance function between the input and the output is:

$$R_{xy}(\tau) = \sum_{k=0}^{\infty} g(kT) R_x(\tau - kT) = g^{\star}(\tau) \star R_x^{\star}(\tau)$$

That is, the cross correlation is the convolution between the impulse response and the autocorrelation of the input. Thus, the impulse response can be estimated from the covariance (correlation if both signals have zero mean) if the input is a white noise. However, this is not a common case. For example, in hydrological systems, we have no control over the time series that always differs from the white noise. This problem is solved by the use of a whitening filter over the input and the output [62, 65].

The so-called cross correlation coefficient (Eq. 2.34) [65] has been widely used to characterize karst systems:

$$r_{xy}(k) = \frac{C_{xy}^N(k)}{\sqrt{\theta_x^2 \theta_y^2}} \tag{2.34}$$

where $C_{xy}^N(k)$ is an estimate of $R_{xy}^N(k)$ for a finite number of samples. Supposing that the input is white noise, the impulsional response of the system can be estimated (Eq. 2.35) as:

$$g(k) = \frac{r_{xy}(k)\sqrt{\theta_y^2}}{\sqrt{\theta_x^2}} \tag{2.35}$$

g(k) is a good estimator only if the input behaves as white noise.

2.7.1.2 The Impulse Response

When the system has a finite impulse response, the non-parametric identification is performed by an intermediate approach between non-parametric and parametric identifications, and corresponds with a representation of the system by a FIR structure. Two methods have been proposed to calculate the coefficients of the impulse response:

Method 1 is based on the Wiener-Hopf summation equation [66, 67]. The estimator of cross-correlation between the input x(t) and output y(t) of the system for an infinite time series, is given by:

$$R_{xy}^N(i) = \sum_{k=0}^{M-1} g(k) R_x(i-k) \quad \text{with } i = 0 \ldots M-1 \tag{2.36}$$

Deploying the Eq. 2.36 in a matrix form [67]:

$$\begin{pmatrix} R_x(0) & R_x(-1) & \cdots & R_x(1-M) \\ R_x(1) & R_x(0) & R_x(-1) & \cdots \\ \cdots & R_x(1) & R_x(0) & R_x(-1) \\ R_x(M-1) & \cdots & R_x(1) & R_x(0) \end{pmatrix} \cdot \begin{pmatrix} g(0) \\ g(1) \\ \cdots \\ g(M-1) \end{pmatrix} = \begin{pmatrix} R_{xy}(0) \\ R_{xy}(1) \\ \cdots \\ R_{xy}(M-1) \end{pmatrix} \tag{2.37}$$

Thus, the impulse response can be calculated as:

$$\mathbf{R}_x \cdot \mathbf{g} = \mathbf{R}_{xy} \Leftrightarrow \mathbf{g} = \mathbf{R}_x^{-1} \cdot \mathbf{R}_{xy} \tag{2.38}$$

Method 2 is based on a minimization process. The most commonly used are linear programming [68, 69] and least squares [70–72]. To identify the impulse response is to minimize the prediction error. Considering a FIR model of length M:

$$Q^\star(t) = \sum_{k=0}^{M-1} g(kT)x^\star(t-kT) + \epsilon(t) = \hat{Q}^\star(t) + \epsilon(t) \tag{2.39}$$

The minimization of $\epsilon(t)$ can be done by several methods. In karst hydrology, least squares, linear programming over the slack variables have been used.

The minimization of prediction error of a FIR model is a case included in the so called parametric identification methods [62]. However, the parametric identification considers also several model structures with infinite impulse response and several models for error behavior.

2.7.2 Non-parametric Identification in Frequency Domain

2.7.2.1 Spectral Analysis

Frequency response can be derived from Fourier transform of the impulse response signal. It provides information about the gain and phase of the system for different input frequencies. The cross spectrum can be calculated in the form:

$$\phi_{xy}^N(n) = \sum_{k=-M+1}^{M-1} w(k) R_{xy}^N(k) e^{-jn\frac{2\pi}{N}k} \tag{2.40}$$

Let $g_T^\star(t)$ the impulse response of a sampled system. Then, as it has been stated, the output $y^\star(t)$ from an input $x^\star(t)$, follows:

$$y^\star(t) = \sum_{k=0}^{\infty} g_T(kT)x^\star(t-kT) + v^\star(t) \tag{2.41}$$

where $v^{\star}(t)$ is the noise in the system. If $v^{\star}(t)$ and $x^{\star}(t)$ are uncorrelated, the (cross) covariance function between the input and the output is given by Eq. 2.34. Supposing the sequences of finite length N, applying the Fourier transform and the equations Eq. 2.34 the result is:

$$\phi_{xy}^{N}(\omega) = G^{N}(j\omega)\phi_{x}^{N}(\omega) \tag{2.42}$$

So, the cross-spectrum is defined as the Fourier transform of the cross-covariance. Thus, an estimate of the frequential transfer function can be obtained as

$$G_{xy}^{N}(j\omega) = \frac{\phi_{xy}^{N}(\omega)}{\phi_{x}^{N}(\omega)} \tag{2.43}$$

However, the result is a plot which is not useful directly for simulation purposes [3].

Another useful function derived from the cross spectrum is the Coherence Function, which is calculated by the following expression:

$$\gamma_{xy}^{2}(\omega) = \frac{|\phi_{xy}^{N}(\omega)|^{2}}{\phi_{x}^{N}(\omega)\phi_{y}^{N}(\omega)} \tag{2.44}$$

It measures the linear correlation between the input and the output of the system at each frequency ω. Notice also that the Coherence Function is dimensionless and can be shown that $0 \leq \gamma_{xy}^{2}(\omega) \leq 1$.

$G_{xy}^{N}(j\omega)$ characterizes the frequential behavior and the coherence $\gamma_{xy}^{2}(\omega)$ characterizes the linearity of a system.

2.7.2.2 The Cross Wavelet Spectrum (XWS)

The bivariate extension of wavelet analysis is recommended when the system involves two time series, instead of only one, to assess time-varying spectral relations between two signals which are often non-stationary.

Cross-wavelet transform (XWS)—also called *wavelet cross-scalogram* or *coscalogram*—gives information on the dependence between two signals, u(t_i) and y(t_i), as a function of time, similar to *cross-correlogram* (or the *cross-spectrogram*). XWS, a bivariate extension of WPS, is commonly applied in earth sciences, e.g., in the experimental cases of this document, the air temperature and glacier discharge or rainfall and conductivity in an aquifer system. The cross wavelet spectrum, introduced by Hudgins et al. [73] to study the atmosphere turbulence, reveals how regions in the time frequency space with large common power have a consistent relationship. This fact suggests a causality between both time series, and the expectation value of the correlation of two signals u(t) and y(t) is:

$$\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\} = \mathscr{W}^{*}_{\psi,s,\tau}\{u(t)\} * \mathscr{W}_{\psi,s,\tau}\{y(t)\} \tag{2.45}$$

Since the wavelet function $\psi(\tau)$ is in general complex, $\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\}$ is also complex.

The cross wavelet power of two signals describes the covariance between these times series at each scale of frequency. Cross wavelet spectrum illustrates quantitatively the similarity of power between two times. It has already been applied to rainfall-runoff cross analysis by Labat et al. [74], and was briefly discussed by Maraun and Kurths [75].

If complex wavelets are used, such as the Morlet wavelet, its squared absolute value $|\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\}|^2$ or simply its absolute value $|\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\}|$ is often gotten for better plotting. The value of $|\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\}|^2$ is large when $|\mathscr{P}_{\psi,s,\tau}\{u(t)\}|$ and $|\mathscr{P}_{\psi,s,\tau}\{y(t)\}|$ are big at the closeness of scales (frequencies) and around the same time, regardless of the local phase difference. When the phase information is required, the Eq. (2.45) should be expressed in terms of its module and phase angle:

$$\begin{aligned}\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\} &= |u(s,\tau)|\, e^{-i\theta_u(s,\tau)}\, |y(s,\tau)|\, e^{i\theta_y(s,\tau)} \\ &= \left|\mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\}\right| e^{i\theta_y(s,\tau)-i\theta_u(s,\tau)}\end{aligned} \tag{2.46}$$

This means that the phase angle $i\theta_y(s,\tau) - i\theta_u(s,\tau)$ reflects the phase difference by which y(t) leads u(t) at the given scale and time. Van Milligen et al. [76] introduced *delayed wavelet cross spectrum*, a useful quantity to detect structures from two separated observation points.

In hydrology, wavelet cross-correlation should sometimes be preferred to classical cross-correlation, because this method provides new insights into the scale depending on the degree of correlation between two given signals [77, 78].

2.7.2.3 Wavelet Coherence Spectrum (CWS)

An extension to Fourier analysis, to allow for non-stationarity, is windowed Fourier analysis, but this overcomes the assumption of global stationarity within each segment. In the time-scale domain, cross-spectrum cannot be normalized locally assuming stationarity to have a value bounded by, for example, zero and one, because multiple points should be involved for some degree of smoothing.

Coherence in signal processing consists of a measure of the correlation between two signals. Power Spectrum represents the power carried by each frequency in a signal. The similarity of two signals can be checked by estimating CWS. Two examples coherence application are: the *wavelet local correlation coefficients* introduced by Buresti and Lombardi [79] to measure the phase coherence of the signals; and *cross wavelet coherence function* introduced by Sello and Bellazzini [80] to assess the intensity coherence of turbulent signals. In practice, to calculate the coherence of two signals, we calculate the cross-power spectrum, because the coherence is the normalized measurement of the cross-power spectrum and can be calculated dividing cross-power spectrum by the squared root of the product of the spectra of the signals

(Eq. 2.47). The expression of wavelet coherence is [81]:

$$\mathscr{C}_{\psi,s,\tau}\{u(t), y(t)\} = \frac{\left\langle \left| \mathscr{X}_{\psi,s,\tau}\{u(t), y(t)\} \right| \right\rangle}{\left(\left\langle \mathscr{P}_{\psi,s,\tau}\{u(t)\} \right\rangle \cdot \left\langle \mathscr{P}_{\psi,s,\tau}\{y(t)\} \right\rangle \right)^{1/2}} \tag{2.47}$$

The $\langle . \rangle$ operator was suggested by Torrence and Compo [43] as the best compromise solution, providing the minimum amount of smoothing necessary to include two independent points in time and scale. Polar expression is a useful form of wavelet coherence with modulus $\rho_{\psi,s,\tau}$ and phase $\phi_{\psi,s,\tau}$, such as:

$$\begin{aligned} \mathscr{C}_{\psi,s,\tau}\{u(t), y(t)\} &= \rho_{\psi,s,\tau}\{u(t), y(t)\} \cdot e^{i\phi_{\psi,s,\tau}(u(t),y(t))} \\ \phi_{\psi,s,\tau} &= \tan^{-1}\left[\mathrm{Im}\left(\mathscr{C}_{\psi,s,\tau}\right) / \mathrm{Re}\left(\mathscr{C}_{\psi,s,\tau}\right) \right] \end{aligned} \tag{2.48}$$

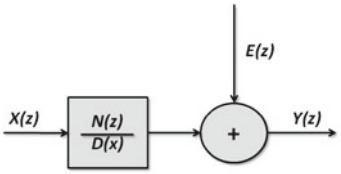
A value of 1 means a perfect coupling between u(t) and y(t) around time τ on a scale s for the wavelet ψ. For a zero or negative value the variations of two signals are not correlated, and for positive value between zero and one the variations are correlated in a certain degree. This value meaning is similar to a traditional correlation coefficient as defined by Barret [82], and it is useful to think of the wavelet coherence as a localized correlation coefficient in time frequency space. Due to these properties, wavelet coherence is an increasingly popular method in analyzing hydrological correlated events.

The fact a signal is correlated with other, not only means that some energy in a frequency is present in both signals, but that plus the frequency which is present in both signals is also related by phase.

The magnitude squared coherence is also used, which is the squared value of the cross-power spectrum divided by the product of the power of the spectra of both signals. The measure of coupling average in the scale (frequency) domain between input and output signals u(t) and y(t) would provide information about the relationship of the signals. This can be got by the wavelet cross spectrum coefficient or simply wavelet coherence spectrum (CWS), obtained from the normalized wavelet cross-spectrum (to have values between zero and one).

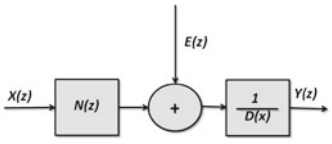
2.8 Parametric Identification

Parametric identification relies on a model previously defined by a set of parameters that must be calculated to accomplish a given quality criteria. E.g., the system characteristics can have a parametric representation through a polynomial of a finite and known degree. The model structure can be obtained by physical modeling (grey box) or it may be a standard one (black box). In the latter case, a set of generic standard structures must be taken into account (OE, FIR, ARX, ARMAX and BJ).

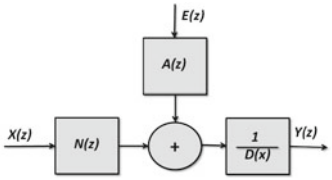


$$OE: Y(z) = \frac{N(z)}{D(z)}X(z) + E(z)$$

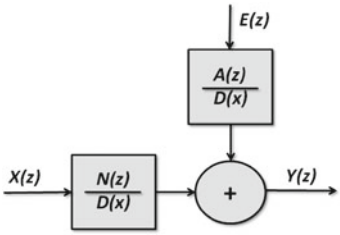
FIR : when $D\,(z) = 1$



$$ARX: Y(z) = \frac{N(z)}{D(z)}X(z) + \frac{1}{D(z)}E(z)$$



$$ARMAX: Y\,(z) = \frac{N(z)}{D(z)}X(z) + \frac{A(z)}{D(z)}E(z)$$



$$BJ: Y(z) = \frac{N(z)}{D(z)}X(z) + \frac{A(z)}{B(z)}E(z)$$

Fig. 2.6 System model structures

2.8.1 Linear Parametric Identification

Parametric identification techniques depend mostly on Prediction-Error Methods (PEM) [62]. The output of system $y^{\star}(t)$ can be expressed as Eq. 2.33. A more useful expression is based on the Z transform:

$$Y(z) = G(z)X(z) + W(z) \tag{2.49}$$

This expression can be rewritten as:

$$Y(z) = G(z)X(z) + H(z)E(z) = \frac{N(z)}{D(z)}X(z) + \frac{A(z)}{B(z)}E(z) \tag{2.50}$$

where $E(z)$ is the transform of a white noise, $\epsilon(t)$. $G(z)$, the transfer function of the system, and $H(z)$, the stochastic behavior of noise. $G(z)$ and $H(z)$ are rational functions whose numerator and denominator are polynomials of the z variable. The relationship between both functions defines several model structures. Figure 2.6 shows the most common ones: AutoRegressive eXogeneous (ARX) model, AutoRegressive Moving Average eXogeneous (ARMAX) model, Box-Jenkins (BJ) model and Output Error (OE) models. The features and advantages of each structure have been studied in several works, see for example Ljung [62].

The ARX model D(z)Y(z) = N(z)X(z) + E(z) is the easiest to estimate since the corresponding estimation problem is of a linear regression type. The foremost

disadvantage is that the disturbance model 1/N(z) comes along with the system's poles. It is, therefore, easy to get an incorrect estimate of the system dynamics because the A (Eq. 2.50) polynomial can also include the disturbance properties. So, higher orders in A and B coefficients in the Eq. 2.50 may be required. If the signal-to-noise ratio is good, this disadvantage is less important.

The ARMAX model $D(z)Y(z) = N(z)X(z) + A(z)E(z)$ has more flexibility in the handling of disturbance modeling than the ARX model. For this reason, ARMAX is a widespread used model and performs well in many engineering applications.

The OE model $Y(z) = [N(z)/D(z)]u(z) + E(z)$ has the advantage that the system dynamics can be described separately and that no parameters are wasted on a disturbance model. If the system operates without feedback during the data collecting, a correct description of the transfer function $G(z) = N(z)/D(z)$ can be obtained regardless of the nature of the disturbance.

In the BJ model $Y(z) = [N(z)/D(z)]u(z) + [A(z)/B(z)]E(z)$ the disturbance properties are modeled separately from the system dynamics.

2.8.2 Selection and Verification Criteria

To get a model reliable, the results and predictions inferred from model should be verified and validated. Model validation is carried out by comparing the model behavior with the system's one and evaluating the difference. All models have a certain domain of validity. This may determine how exactly they are able to describe the system behavior.

Efficiency criteria can be defined as mathematical measures of how well a model simulation fits the available observations [83]. A number of different methods to set a criterion have been suggested in the literature, e.g., least squares, generalized least squares, maximum likelihood or instrumental variables. Krause et al. [84] have studied the utility of several efficiency criteria in three examples using a simple observed streamflow hydrograph, declaring that: "The selection and use of a specific efficiency criteria and the interpretation of the results can be a challenge for even the most experienced hydrologist, since each criterion may place different emphasis on different types of simulated and observed behaviors".

Nash-Sutcliffe efficiency, coefficient of determination, and index of agreement, are frequently applied to verify hydrologic models. The efficiency value E proposed by Nash and Sutcliffe [85] is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period (Eq. 2.51).

$$E = 1 - \frac{\sum_{k=1}^{N} \left[y(kT) - \hat{y}(kt) \right]^2}{\sum_{k=1}^{N} \left[y(kT) - \overline{y} \right]^2} \tag{2.51}$$

where $y(kT)$ is observed data in the sampling interval T (with k $= 0, 1, 2, \ldots,$ N), $\hat{y}(kt)$ is modeled output, and $\overline{y}$ is mean of observed data. Nash-Sutcliffe efficiencies can range from $-\infty$ to 1. An efficiency value of 1 (E $= 1$) corresponds to a perfect match of model output to the measured data. An efficiency value of 0 (E $= 0$) indicates that the model is as accurate as the mean of the observed data, whereas an efficiency less than zero (E < 0) occurs when the observed mean is a better predictor than the model or, in other words, when the residual variance (described by the numerator in the expression above), is larger than the data variance (described by the denominator). Essentially, the closer the model efficiency value is to 1, the more accurate the model is.

According to Legates and McCabe [86], the disadvantage of the Nash-Sutcliffe efficiency is the fact that the differences between the observed and predicted values are calculated as squared values. As a result, larger values in a time series are strongly overestimated, whereas lower values are neglected. E.g., in the quantification of an aquifer discharge, this criterion could lead to an overestimation of the model performance during peak flows and an underestimation during low flow conditions. Nevertheless, Nash-Sutcliffe efficiency, expressed frequently as 0–1 coefficient (E), is the unique criterion used in the experimental cases of this thesis, in order to standardize the comparison between models.

2.9 Nonlinear Identification

It is difficult to establish a clear identification methodology of nonlinear systems, since analysis is usually more intricate than in the identification of linear models, because of the variety of nonlinear model structures and nonlinear behaviors.

Donoho and Johnstone [87], and Donoho [88] introduced nonlinear wavelet estimators in nonparametric regression through thresholding, i.e., the term-by-term assessment of coefficients in the wavelet expansion. Only coefficients that exceed a predetermined threshold are taken in account. This produces the wavelet shrinkage. Bendat [89] describes procedures to identify and analyze the properties of many types of nonlinear systems as Zero-Memory Nonlinear Systems and Parallel Nonlinear System, with analysis of Nonlinear System Input/Output Relationships. Zhang [90] applied wavelet theory for nonlinear system identification, with a wavelet basis as a universal function approximator, with a neural network used to determine the resolution, and the translation coefficients of the wavelet. This nonparametric estimator named wavelet network has a neural network like structure that makes use of techniques of regressor selection completed with backpropagation procedure.

This section is going to focus only on nonlinear parametric identification, and, inside this type, Volterra series and Hammerstein- Wiener methods.

2.9.1 Nonlinear Parametric Identification

2.9.1.1 Volterra Identification

Volterra series have been widely applied as nonlinear system modeling technique. These Volterra representations lead to very complicated identification algorithms since evaluation of higher order Volterra kernel often require very large amounts of data for even low order nonlinearities. However, kernels can be calculated for systems whose order is known and finite.

Some identification procedures to calculate higher-order kernels have given by [91], but the proposed solutions reduce their practical application only to model system of low non-linearity. Mirri et al. [92] suggested three different models for Volterra series, in order to reduce the number of kernels that are considered in the identification process of a non-linear system. The reduced number of mathematical operators involved, simplifies the experimental procedures, but it needs a somehow information about the system behavior.

2.9.1.2 Hammerstein-Wiener Identification

The literature about the Hammerstein-Wiener model identification is ample [93, 94]. Most important methods try to reduce the parameter redundancy by using linear and nonlinear structures. Other methods use optimized algorithms to decrease the computation complexity.

- The iterative algorithm, a classical method proposed by Narendra and Gallman [95], parametrizes the system for the linear prediction error. In the most common version of this method, the parameter set is usually divided into two subsets. One set is fixed while the other searches the optimal values. Then, both sets are switched to perform the reverse operation. The estimation is carried out by minimizing alternatively with respect to each set of parameters, a quadratic criterion on the prediction errors. A main problem of the iterative algorithm is the convergence.
- Bai et al. [96] studied the convergence properties of iterative algorithm from Narendra and Gallman [95] They show that the iterative algorithm with normalization is convergent in general, and for finite-impulse response, the convergence is reached in one step.
- A noniterative method was proposed by Chang and Luus [97] and show through numerical examples that the computation time by this method is considerably less than by the iterative, while the accuracy of the estimates is comparable.
- Bai [98] presented an optimal two stage identification algorithm for Hammerstein-Wiener model. The first step is the recursive least squares. The second one is the singular value decomposition of two matrices with fixed dimensions and do not increase as the number of the data point increases.
- Reference [99] described a blind approach, where the linear part is only using the output measurements, i.e., no information on the input.

- Geothals et al. [100] extend the unifying theorem for three subspace system identification algorithms by using Least Squares Support Vector Machines component to identify Hammerstein-Wiener structure.
- Wills et al. [94] illustrates a new maximum-likelihood based method for the identification of Hammerstein-Wiener model structures.

2.10 Conclusion

To study system dynamics, some stages are recommended: Preparation, analysis, model structures preselection and identification.

Raw data need refining and adjustment before time series analysis because of outliers or abnormal values in the readings. Among other criteria, Wavelet-Rosner is an efficient test for outlier detection in the frequency domain. Also, the selection of the proper sampling frequency is discussed.

In analysis under the scenario of the time domain, autocorrelation is a lag correlation of a given time series within itself, lagged by a number of times units. Positive autocorrelation is a specific form of 'persistence' of events, and time series is more predictable. The wavelet transform calculates the correlation between the signal and a wavelet function $\psi(t)$, the mother wavelet, parametrized by location and scale. The similarity between the signal and the wavelet function is computed separately for different time intervals, resulting in a three dimensional representation. In the frequency domain analysis, the power density spectrum assesses the energy level of the signal in different frequency bands.

Linear time-invariant models provide structures for systems. Although in nature all systems are nonlinear, in practical cases the use of linear models is justified in order to simplify the study or as a reference for other more exhaustive models. The effect of any invariant linear system (LTI) on an arbitrary input signal is obtained by convolution of the input signal with the system's impulse response function. A Transfer Function in LTI can be expressed as the ratio of the Laplace transform of the output and the input, and corresponds to the Laplace transform of the impulse response. The frequential transfer function characterizes the frequential behavior of the system, that is, how the frequential components of the input are modified (amplitude change and phase delay) to compose the output.

The non-linear models by Volterra equations are not suitable in the case of long uninterrupted records at a finer sampling rate, because Volterra models have a stationary or time invariant kernel. Nevertheless, block-oriented models have emerged as an appealing proposal due to their simplicity and their property of being valid over a larger operating region than a LTI model and are easier to implement than heavy-duty nonlinear models, such as neural networks and Volterra models.

Nonparametric identification techniques provide a very effective and simple way of finding model structure in data sets without the imposition of setting required parameters. Cross correlation is the convolution between the impulse response and the autocorrelation of the input. Thus, the impulse response can be estimated from the

covariance if the input is a white noise. When the system has a finite impulse response, the non-parametric identification is performed by an intermediate approach between non-parametric and parametric identifications, and corresponds with a representation of the system by a FIR structure. A nonparametric identification in the frequency domain can be carried out by cross wavelet spectrum. This reveals how regions in the time frequency space with large common power spectrum suggest a causality between both time series. The coherence is the normalized version of cross wavelet spectrum.

Characterizing all the input-output properties of a system through exhaustive measurements is usually impossible. Instead, parametric identification is a way to make that a finite number of measurements help infer the system response to specified inputs. Parametric identification techniques, which depend mostly on Prediction-Error Methods (t), can be expressed by Z transform. Model validation is carried out by comparing the data generated by the model with those observed in the system. Nash-Sutcliffe efficiency, coefficient of determination, and index of agreement, are frequently applied to verify hydrologic models.

It is difficult to establish a clear identification methodology of nonlinear system identification, because of the variety of nonlinear model structures and nonlinear behaviors. These Volterra representations lead to very complicated identification algorithms since evaluation of higher order Volterra kernels often requires very large amounts of data for even low order nonlinearities. However, kernels can be calculated for systems whose order is known and finite. Most important methods in Hammerstein-Wiener model identification try to reduce the parameter redundancy by using linear and nonlinear structures. Other methods use optimized algorithms to decrease the computation complexity.

References

1. Williams T (1988) Modelling complex projects. Wiley, New Jersey
2. Eykhoff P (1974) System identification. Wiley, London
3. Ljung L, Glad T (1994) Modeling of dynamic systems. Prentice Hall, Englewood Cliffs
4. Alippi C, Anastasi G, Di-Francesco M, Roveri M (2009) Energy management in wireless sensor networks with energy-hungry sensors. IEEE Instrum Meas Mag 12(2):16–23
5. Hernàindez L, Baladrón C, Aguiar JM, Calavia L, Carro B, Sàinchez-Esguevillas A, Cook DJ, Chinarro D, Gómez J (2012) A study of the relationship between weather variables and electric power demand inside a smart grid/ smart world framework. Sensors 12:11571–11591
6. Zayed AI (1993) Advances in shannon's sampling theory [Hardcover]. CRC Press, Library of Congress, USA
7. Hawkins D (1980) Identification of outliers. Chapman and Hall, London
8. Barnett V, Lewis T (1994) Outliers in statistical data. Wiley, Chichester
9. Tabachnick B, Fidell L (1996) Using multivariate statistics. Harper-Collins College Publishers, New York
10. Rosner B (1983) Percentage points for a generalized esd many-outlier procedure. Technometrics 25:165–172
11. Chinarro D, Villarroel J, Cuchí J (2011) Wavelet analysis of fuenmayor karst spring, San Julián de Banzo, Huesca, Spain. Environ Earth Sci Spec Issue 65(8):1–13

12. Tukey J (1960) A survey of sampling from contaminated distributions. In: Olkin I et al (eds) contributions to probability and statistics: essays in honor of harold hotelling. Stanford University Press, Stanford, pp 448–485
13. Box G, Jenkins G (1976) Time series analysis: forecasting and control. Holden-Day, San Francisco
14. Stokes G (1879) Note on searching for hidden periodicities. Proc R Soc 29:122–125
15. Schuster A (1898) On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terr Magn Atmos Electricity (J Geophys Res) 3(1):13–41
16. Wellstead P (1981) Non-parametric methods of system identification. Automatica 17(1): 55–69
17. Welch P (1967) The use of the fast fourier transform for the estimation of the power spectra: a methods based on time averaging over short modified periopdograms. IEEE Trans Audio Electroacustic 15(2):70–73
18. Ljung L, Glover K (1981) Frequency domain versus time domain methods in system identification. Automatica 17(1):71–86
19. Blackman R, Tukey J (1958) The measurement of power spectra. Dover Publications, Inc., New York
20. Grossman A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. J Math Anal 15:732–736
21. Gabor D (1946) Theory of communication. J IEE 93(26):429–457
22. Kronland-Martinet R, Morlet J, Grossmann A (1987) Analysis of sound patterns through wavelet transforms. J Pattern Recognit 1(2):273–302
23. Haar A (1910) Zur theorie der orthogonalen funktionensysteme. Math Ann 69:331–371
24. Esteban D, Galand C (1977) Application of quadrature mirror filters to split-band voice coding schemes. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP-77, vol 2. pp. 191–195
25. Crochiere RE, Webber SA, Flanagan JL (1976) Digital coding of speech in subbands. Bell Syst Tech 85:1069–1085
26. Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Proc Geophys 11:561–566
27. Sangbae Kima FI (2007) On the relationship between changes in stock prices and bond yields in the G7 countries: wavelet analysis. J Int Finan Markets, Inst Money 17:167–179
28. Ricker DW (2003) Echo signal processing (The springer international series in engineering and computer science). Kindle Edition
29. Rivera D, Lillo M, Arumí J (2007) ENSO influence on precipitation of chillan, chile: an approach using wavelets. Gestión Ambiental 13:33–48
30. Meyers SD, Kelly BG, O'Brien JJ (1992) An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of yanai waves. Mon Weather Rev 121(13–31):101–105
31. Cuchí J, Villarroel J (2002) Análisis de respuesta de Fuenmayor (San Julián de Banzo, Huesca). Primeros resultados. Geogaceta 31:75–78
32. Labat D, Ababou R, Mangin A (1999a) Analyse en ondelettes en hydrologie karstique. Geoscience 329:881–887
33. Hernández L, Baladrón C, Aguiar JM, Carro B, Sánchez-Esguevillas A, Lloret J, Chinarro D, Gómez JJ, Cook D (2013) A multi-agent-system architecture for smart grid management and forecasting of energy demand in virtual power plants. IEEE Commun 51(1):106–114
34. Ebadi L, Shafri HZM, Mansor SB, Ashurov R (2013) A review of applying second-generation wavelets for noise removal from remote sensing data. Environ Earth Sci 70(6):2679–2690. doi:10.1007/s12665-013-2325-z
35. Amat S, Muñoz J (2008) Some wavelets tools for maxwell's equations. Carthagonova J Math 1(1):1–18
36. Daubechies I, Daniel Kleppner SM, Meyer Y, Ruskai MB, Weiss G (2001) Wavelets: seeing the forest and the trees. Office on Public Understanding of Science,Technical report, National Academy of Sciences

37. Daubechies I (1990) The wavelet transform time-frequency localization and signal analysis. IEEE Trans Inf Theory 36(5):961–1004
38. Foufoula-Gergiou E, Kumar P (1994) Wavelets in geophysics, (wavelet analysis and its applications), 1st edn. Academic Press, Boston
39. Heil CE (1989) Continuous and discrete wavelet transforms. SIAM Rev 31(4):628–666
40. Mallat S (1998) A wavelet tour of signal processing, 2nd edn. Academic Press, École Polytechnique, Paris, Courant Institute, New York University
41. Struzik Z, Siebes A (2000) Outlier detection and localisation with wavelet based multifractal formalism. Inf Syst (INS), R0008
42. Mallat S, Zhong S (1992) Characterization of signals from multiscale edges. IEEE Trans Pattern Anal Mach Intell 14(7):710–732
43. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bulletin Am Meteorol Soc 79:61–78
44. Farge (1992) Wavelet transforms and their applications to turbulence. Annu Rev Fluid Mech 24:395–457
45. Strang G (1993) Wavelet transforms versus fourier transforms. Bulletin, new series, Am Math Soc 28:288–305
46. Perrier V, Philipovitch T, Basdevant C (1995) Wavelet spectra compared to fourier spectra. J Math Phys 36:1506–1519
47. Popivanov I, Miller RJ (2001) Similarity search over time series data using wavelets. Technical report TR CSRG 438. Department of Computer Science, University of Toronto, 438
48. Becher V, Bienvenu L, Downey R, Mayordomo E (2012) Computability, complexity and randomness (dagstuhl seminar 12021). Dagstuhl Rep 2(1):19–38
49. Haber R (1985) Nonlinearity test for dynamic processes. IFAC Identification and system parameter estimation symposium. Pergamon Press, Elmsford, New York, pp 409–414
50. Giannakis GB, Serpedin E (2001) A bibliography of non linear system identification. Signal Process 81:553–580
51. Wiener N (1958) Nonlinear problems in random theory. MIT Press, Cambridge
52. Singleton HE (1950) Theory of nonlinear transducers. Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Technical Report 160
53. Bose AG (1959) A theory of nonlinear systems. Technical report, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge
54. Brilliant MB (1958) Theory of the analysis of nonlinear systems. Technical report, 345. Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge
55. George DA (1959) Continuous nonlinear systems. Technical report, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge
56. Minorsky N (1947) Introduction to nonlinear mechanics. J. W. Edwards Press, Ann Arbor, Michigan
57. Schetzen M (1980) The volterra and wiener theories of nonlinear systems. Wiley, New York
58. Leontaritis I, Billings S (1985) Input-output parametric models for nonlinear systems. part i-deterministic non-linear systems. Int J Control 41:303–328
59. Pottmann M, Pearson RK (1998) Block-oriented NARMAX models with output multiplicities. AIChE J 44:131–140
60. Eskinat E, Johnson SH, Luyben WL (1991) Use of Hammerstein models in identification of nonlinear systems. AIChE J 37:255–268
61. Wigren T (1993) Recursive prediction error identification using the nonlinear wiener model. Automatica 29(4):1011–1025
62. Ljung L (1999) System identification. theory for the user., 2nd edn. Prentice Hall, New Jersey
63. Deistler M (2002) System identification and time series analysis: Past, present, and future. In: Pasik-Duncan B (ed) Stochastic theory and control, lecture notes in control and information sciences, vol 280. Springer, Berlin Heidelberg, pp 97–109
64. Aström K, Bohlin T (1965) Numerical identification of linear dynamic systems from normal operating records. In Proceedings of IFAC symposium on self-adaptive systems, Teddington, UK

65. Box G, Jenkins G, Reinsel G (1994) Time Series Analysis. Forecasting and Control, 3rd edn. Prentice Hall, Englewood Cliffs
66. Labat D, Ababou R, Mangin A (1999) Analyse en ondelettes en hydrologie karstique. 2e partie: Analyse multirésolution croissée de pluie-débit de sources karstiques. Comptes Rendus de Académie des Sciences. Geoscience. 329:881–887
67. Labat D, Ababou R, Mangin A (2000a) Rainfall-runoff relations for karstic springs. part I: convolution and spectral analyses. J Hydrol 238:123–148
68. Dreiss S (1982) Linear kernels for karst aquifers. Water Resour Res 18(4):865–876
69. Dreiss S (1983) Linear unit-response functions as indicators of recharge areas for large karst springs. J Hydrol 61:31–44
70. Dreiss S (1989) Regional scale transport in a karst acuifer. 2. linear systems and time moment analysis. Water Resour Res 25(1):126–134
71. Padilla A, Pulido-Bosch A (1990) A reservoir model to simulate karstic aquifers. In: International conference on calibration and reliability in groundwater modeling, ModelCARE90, pp. 323–333
72. Denić-Jukić V, Jukić D (2003) Composite transfer functions for karts aquifers. J Hydrol 274:80–94
73. Hudgins L, Friebe C, Mayer M (1993) Wavelet transforms and atmospheric turbulence. Phys Rev Lett 71(20):3279–3282
74. Labat D, Ababou R, Mangin A (2001) Nonlinearity and nonstationarity in nainfall-runoff relations for karstic springs. Institut de Mécanique des Fluides de Toulouse. Laboratoire Souterrain de Moulis. The International Association of Hydraulic Engineering and Research
75. Maraun D, Kurths J (2004) Nonlinear analysis of multivariate geoscientific data - advanced methods, theory and application. Department of Physics, Potsdam University, D-14415 Potsdam, Germany
76. Van Milligen B, Sànchez E, Estrada T, Hidalgo C (1995) Wavelet bicoherence: a new turbulence analysis tool. Physics Plasmas 8:3017
77. Onorato M, Salvetti M, Buresti G, Petagna P (1997) Application of a wavelet cross-correlation analysis to dns velocity signals. Eur J Mech 16(4):575–597
78. Labat D, Ababou R, Mangin A (2000b) Rainfall-runoff relations for karstic springs. part ii: continuous wavelet and discrete orthogonal multiresolution analyses. J Hydrol 238:149–178
79. Buresti G, Lombardi G (1999) Application of continuous wavelet transforms to the analysis of experimental turbulent velocity signals. In: Banerjee S, Eaton JK (eds) Turbulence and Shear Flow Phenomena, Begell House Inc, pp. 762–772
80. Sello S, Bellazzini J (2000) Wavelet cross-correlation analysis of turbulent mixing from large-eddy simulations. Eighth European Turbulence Conference, Barcelona pp. 27–30
81. Liu PC (1994) Wavelet spectrum analysis and ocean wind waves. In: Foufoula-Georgiou E, Kumar P. (eds.) Wavelets in geophysics pp. 151–166
82. Barrett JF (1964) Hermite functional expansions and the calculation of output autocorrelation and spectrum for any time-invariant nonlinear system with noise. J Electron Control 16: 107–113
83. Beven J (2001) Rainfall-Runoff modelling. the primer. Wiley, Chichester
84. Krause P, Boyle D, Base F (2005) Comparison of different efficiency criteria for hydrological model. Adv Geosci 5:89–97
85. Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models. part i. a discussion of principles. J Hydrol 10:282–290
86. Legates D, McCabe G (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour 35:233–241
87. Donoho D, Johnstone I (1994) Ideal spatial adaptation by wavelet shrinkage. Biomefrika 81:425–455
88. Donoho D (1995) De-Noising by soft-thresholding. IEEE Trans Inf Theory 41(3):613–627
89. Bendat J (1990) Nonlinear system analysis and identification. Willey, New York
90. Zhang Q (1997) Using wavelet network in nonparametric estimation. IEEE Trans Neural Network 8(2):227–236

91. Glentis G, Koukoulas P, Kalouptsidis N (1999) Efficient algorithms for volterra system identification. IEEE Trans Signal Process 47(11):3042–3057
92. Mirri D, Iuculano G, Traverso P, Pasini G, Filicori F (2003) Non-linear dynamic system modelling based on modified volterra series approaches. Measurement 33:9–21
93. Giri F, Bai E (2010) Block-oriented Nonlinear System Identification. In: Control and Information Sciences (Lecture Notes). Springer, Heidelberg
94. Wills A, Schön T, Ljung L, Ninness B (2013) Identification of hammerstein-wiener models. Automatica 49:70–81
95. Narendra KS, Gallman PG (1966) An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Trans Autom Control 11(3):546–550
96. Bai E, Kullback DL, Leibler RA (2004) Convergence of the iterative hammerstein system identification algorithm. IEEE Trans Autom Control 49(11):1929–1940
97. Chang F, Luus R (1971) A noniterative method for identification using hammerstein model. IEEE Trans Control Syst Technol 16(5):464–468
98. Bai E-W (1998) An optimal two stage identification algorithm for hammerstein-wiener nonlinear systems. Automatica 34(3):333–338
99. Bai E (2002) A blind approach to the hammerstein-wiener model identification. Automatica 38:967–979
100. Goethals I, Pelckmans K, Suykens J, De-Moor B (2005) Subspace identification of hammerstein systems using least squares support vector machines. IEEE Trans Autom Control 50(10):1509–1519

http://www.springer.com/978-3-319-08857-0

System Engineering Applied to Fuenmayor Karst Aquifer (San Julián de Banzo, Huesca) and Collins Glacier (King George Island, Antarctica)

Chinarro, D.

2014, XX, 161 p. 54 illus., 33 illus. in color., Hardcover
ISBN: 978-3-319-08857-0