Nonparametric Data-Driven Modeling of Linear Systems

ESTIMATING THE FREQUENCY RESPONSE AND IMPULSE RESPONSE FUNCTION

JOHAN SCHOUKENS, KEITH GODFREY, and MAARTEN SCHOUKENS

Digital Object Identifier 10.1109/MCS.2018.2830080
Date of publication: 18 July 2018

The aim of this article is to give a tutorial overview of frequency response function (FRF) or impulse response (IR) function measurements of linear dynamic systems. These nonparametric system identification methods provide a first view on the dynamics of a system. As discussed in "Summary," the article discusses three main points. The first replaces

1066-033X/18©2018IEEE

Summary

Frequency response function (FRF) and impulse response (IR) measurements provide a first view on the dynamics of a system. Previously developed FRF measurement methods were optimized for the available computer power at that time. These classical methods are still popular in engineering curricula and industry. However, more advanced algorithms are now available that can reduce the measurement time and errors by a factor of two to 100 by making better use of the increased compute power. Because these methods are not well known to the public, they are not frequently used, leading to a waste of money, resources, and time. The goal of this article is to bridge this gap by providing a deeper insight into the underlying problems of FRF and IR measurements and to use this better understanding to introduce the recent more powerful methods. Links to publicly available software are provided, which helps to reconstruct the results shown in this article and minimizes the effort to adopt these new algorithms.

classic FRF measurement techniques based on spectral analysis methods with more advanced, recently developed algorithms. User guidelines will be given to select the best among these methods according to four specific user situations: 1) measurements with a high or low signal-to-noise ratio (SNR), 2) systems with smooth or fast-varying transfer functions as a function of the frequency, 3) batch or real-time processing, and 4) low or high computational cost. The second main point is to store the reference signal together with the data. This will be very useful whenever there are closed loops in the system to be tested, including interactions between the generator and the setup. The final point is to use periodic excitations whenever possible. Periodic excitations provide access to a full nonparametric noise model, even under closed-loop experimental conditions. Combining periodic signals with the advanced methods presented in this article provides access to high-quality FRF measurements, while the measurement time is reduced by eliminating disturbing transient effects.

Depending upon the situation, it might be possible to reduce the measurement time or error by a factor of 2–100.

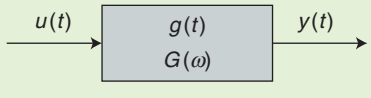


FIGURE 1 A continuous or discrete-time (DT) system is considered with impulse response $g(t)$ and a frequency response function $G(\omega)$. For a continuous-time system, $G(\omega)$ is a shorthand notation for $G(s = j\omega)$, and for a DT system, $G(\omega)$ stands for $G(z = e^{j\omega})$.

The complete palette that was recently developed ranging from simple classical methods to advanced methods will be covered. The article provides a deep understanding of the problems related to FRF and IR measurements. These insights are used as a basis to understand the classical methods and explain how these methods can be improved to obtain better FRF-IR measurements than the classical methods that still dominate the field. This leads to a completely new class of IR-FRF estimators.

FROM THE CLASSIC TIME AND FREQUENCY-DOMAIN METHODS TO RECENT ADVANCED PROCESSING TECHNIQUES

The overview of the classic time- and frequency-domain methods, developed in the 1950s and 1960s, will be complemented by an introduction to the more powerful methods that were developed in the last decade. The measurement time can be significantly reduced using these new techniques, at a cost of increased computational demands. Because the available computational power has grown by several orders of magnitude since the late 1960s, it is clear that there is no reason to continue to use the initial choices from 50 years ago. The computational time was the dominant constraint driving prior research efforts, confining the algorithms to be simple [1]–[4] (see the section "Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods"). However, more complex algorithms are currently possible that will either reduce the measurement time or improve the quality of the measurements. It is the goal of this article to make these new algorithms accessible for a wide group of measurement and control engineers.

FREQUENCY RESPONSE FUNCTION AND IMPULSE RESPONSE FUNCTION MEASUREMENTS

The goal is to obtain the IR or FRF of the system in Figure 1 (together with a confidence bound), starting from discrete-time (DT) measurements of the input $u(kT_s)$ and output $y(kT_s)$, $k = 1,2,\ldots,N$, with T_s the sample period that is the inverse of the sample frequency f_s. The FRF measurement evaluates $G(\omega)$ at a discrete set of frequencies $f_k = kf_0 = kf_s/N$, while the IR measurement returns an estimate for $g(t)$ at a discrete time grid kT_s. Both measurements provide equivalent information. The FRF is the Fourier transform of the IR. Depending upon the needs, some information can be more easily accessed in the time or frequency domain. A dominant resonance can be more easily analyzed in the frequency domain, and the presence of dead time will be more easily observed in the time domain. The choice between them is set by the user's needs, experiences, and personal preferences.

For didactic reasons, a major part of the article is focused on single-input, single-output (SISO) systems. "Frequency Response Function Measurements for MIMO Systems" gives an introduction to the multiple-input, multiple-output

(MIMO) systems while paying special attention to the design of the experiment [5]–[9]. Recent FRF measurements using some of the methods presented in this article are discussed in [10].

The discussion considers both continuous-time (CT) and DT systems. The actual nature of the system does not critically influence the presented algorithms, provided that the sampling frequency f_s is sufficiently high.

TIME-DOMAIN METHODS TO ESTIMATE THE IMPULSE RESPONSE: BASIC IDEAS

Consider the system in Figure 1 with

$$y(t) = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)d\tau = g(t) * u(t), \tag{1}$$

where * denotes the convolution. A first class of methods estimates the IR $g(t)$, starting from the measured input and output signals by solving the deconvolution problem in (1). The equivalent description based on the cross-correlation $R_{yu}(\tau) = E\{y(t+\tau)u(t)\}$ and autocorrelation function $R_{uu}(\tau) = E\{u(t+\tau)u(t)\}$ is very useful in simplifying the computations [4], [11], [12]

$$R_{yu}(\tau) = g(\tau) * R_{uu}(\tau). \tag{2}$$

White random noise excitations reduce the autocorrelation of the input to $R_{uu}(\tau) = \sigma_u^2\delta(\tau)$, and the cross-correlation becomes

$$R_{yu}(\tau) = \sigma_u^2 g(\tau), \tag{3}$$

allowing the IR to be measured directly, without making an explicit deconvolution. In the 1960s, pseudorandom binary sequences (PRBSs) [13], [14] were used to replace the white noise excitations, resulting in a lower uncertainty on the estimated IR for a given measurement time. These methods are discussed in the "Time-Domain Approach" section.

Using the increased computer power, it is currently possible to directly estimate the IR, even for arbitrary excitations. Combining the experimental data with prior user information like exponential decay and smoothness of the IR function further reduces the uncertainty [15]. These aspects are discussed in detail in the section "Variance Reduction by Combining Data and Prior Knowledge: Regularization."

FREQUENCY-DOMAIN METHODS TO ESTIMATE THE FREQUENCY RESPONSE FUNCTION: BASIC IDEAS

An alternative approach to the deconvolution problem in (1) or (2) to estimate $g(t)$ is to transform (1) to the frequency domain. Define $U(k), Y(k)$ as the discrete Fourier transform (DFT) of the measured input and output u,y with k the frequency index (see "The Discrete Fourier Transform"). Neglecting the finite length measurement effects, the following relation holds [16]

$$Y(k) = G(k)U(k). \tag{4}$$

It is very tempting to estimate the FRF by direct division of $Y(k)/U(k)$. However, this only works well if $U(k)$ does not become very small or equal to zero (see the section "Frequency Response Function Measurements Using Periodic Excitations"). This can be realized using well-designed signals (see "Design of Excitation Signals"). However, it is better to average the data over multiple subrecords before the division. To do so, the Fourier transform of (1) is replaced by the Fourier transform of (2), leading to the crossspectrum $S_{yu} = F(R_{yu})$ and autospectrum $S_{uu} = F(R_{uu})$. With $S = F(R)$, the Fourier transform of (2) becomes

$$G(k) = S_{yu}(k)/S_{uu}(k). \tag{5}$$

This method became the standard approach in the 1960s and is still used today in all dynamic signal analyzers [1]–[4], [11], [17], [18] (see the sections "Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods" and "Time and Frequency-Domain Interpretation of Windows").

Recently, a new class of spectral methods was developed that provides superior results over (5) [19], [20], again at the cost of higher computational demands (see the section "Improved Frequency Response Function Measurements Using Local Parametric Methods"). With the computational resources that are currently available, these new methods should be the default choice.

The Discrete Fourier Transform

Consider a discrete time sequence $x(t), t = 0, 1, \ldots, N-1$. The discrete Fourier transform (DFT) and the inverse discrete Fourier transform (IDFT) are then given by

$$X(k) = \frac{1}{\sqrt{N}}\sum_{t=0}^{N-1} x(t)e^{-j2\pi tk/N}, \tag{S1}$$

and

$$x(t) = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} X(k)e^{j2\pi kt/N}, \tag{S2}$$

$X(k)$ is the Fourier coefficient of $x(t)$ at frequency $f_k = kf_s/N$, with $f_s = 1/T_s$ the sampling frequency of the discrete time sequence. The scale factor of the DFT can vary from one definition to the other, depending on the purpose. For random excitation, the $1/\sqrt{N}$ is commonly used because it returns an averaged amplitude that is independent of N.

The overview of the classic time- and frequency-domain methods, developed in the 1950s and 1960s, will be complemented by an introduction to the more powerful methods that were developed in the last decade.

PARAMETRIC AND NONPARAMETRIC MODELS

Parametric models are described by a finite number of parameters, and this number does not depend upon the data length. A typical example is a transfer function model for a linear dynamic system, where the number of parameters is set by the number of poles and zeros. In nonparametric models, the number of parameters grows with the data length. A typical example is the FRF of a system. It will be shown that the frequency resolution is inversely proportional to the data length, and hence the number of frequencies where an FRF measurement is made grows proportional to the data length. There is no sharp definition for both classes of models. IR measurements can be assigned to both classes. The length of the IR that can be estimated from a given data set grows with the length of the data (a nonparametric model characteristic). However, it depends only logarithmically on the data length, so that it becomes almost constant for longer data records (a parametric model characteristic). Usually, it makes no sense to estimate the IR for more than a few time constants of the system.

This article focuses only on nonparametric identification. There is extensive literature on the topic of parametric identification, including [16] and [21]–[23].

EXPERIMENT DESIGN: CHOICE OF THE EXCITATION SIGNAL

In both the time- and the frequency-domain approach, the choice of the excitation signal has an impact on the methods to be used [24], [25]. The variance of the measured impulse or FRF strongly depends on this choice. Initially, only simple excitations like steps, impulses, or sines could be generated. In the late 1950s and 1960s, more advanced periodic binary excitations could be generated using simple hardware [14], [26]–[28]. These signals simultaneously excite multiple frequencies with the same power (see "Design of Excitation Signals" and Figure S2), which was a major step forward to reduce the required measurement time.

In the late 1970s and early 1980s, arbitrary random generators became available, so it then became possible to directly generate advanced signals that were designed and computed on a digital computer. This opened many possibilities to increase the SNR of the measurements and reduce the impact of nonlinear distortions on the FRF measurement [16], [29], [30]. "Design of Excitation Signals" provides a detailed discussion of these signals, including the relevant design parameters, user choices, and advantages and disadvantages.

OUTLINE

The article is organized as follows. First, the measurement setup and noise assumptions are presented. The FRF methods are split along the use of periodic excitations and random excitations. These sections cover both the classical and the recently developed methods. The time-domain methods are also organized in a similar manner. A small historical overview of the methods used before the early 1960s is included, which discusses the elegance of the methods that solved IR and FRF measurement problems. Throughout the article, user guidelines are included that provide practical tips and highlight useful information.

MEASUREMENT SETUP

Each data-driven modeling process should start with a careful inspection of the measurement setup. A short discussion with the plant operators or the measurement team can save significant time. It is very important to know what preprocessing is applied to the data, including if filters were turned on, if drift removal was applied to the raw data, if measurements were obtained around a given set point, if the mean values were included in the raw data, and how outliers and missing data were handled. Without being aware of these actions, significant time and effort might be wasted by including their effects in the model.

Although the setup in Figure 2 covers many interesting situations, it still does not address all realistic ones. Often, the plant to be modeled is a part of a larger complex network with many interacting loops. Under these conditions, it is uncertain if it is possible to isolate the subsystem of interest from the rest of the plant using the available measurements. In [31], [32], and succeeding work, a detailed analysis of the minimum required measurement conditions is made to ensure that the IR/FRF of the actual subsystem is measured.

Intersample Assumptions

All data processing in this article starts from DT measurements $u(kT_s), y(kT_s)$, sampled at a sampling frequency $f_s = 1/T_s$. No information is available on how the CT signals $u(t), y(t)$ vary between the measured samples. For this reason, assumptions are needed, and the measurement setup should match the intersample assumption well. The

Parametric models are described by a finite number of parameters, and this number does not depend upon the data length.

two most popular intersample assumptions [16], [33] are shown in Figure 3: the zero-order hold (ZOH) and the band-limited (BL) assumption. A detailed discussion of both intersample assumptions (the facts and their appreciation) is given in [16, Sec. 13.2 and 13.3].

Zero-Order Hold Setup

The ZOH setup places a condition on the excitation that is assumed to remain constant in between the samples. In this setup, the IR and FRF are estimated between the DT reference signal in the memory of the generator and the sampled output. The intersample behavior and the actuator characteristic are an intrinsic part of the model: if the intersample behavior changes, the corresponding model will also change. The ZOH assumption is very popular in digital control. The sampling frequency f_s is commonly chosen to be ten times larger than the frequency band of interest. A DT model provides an exact description of the CT system.

Band-Limited Setup

The BL setup assumes that above $f_{\max}$ there is no power in the signals: $U(|f| > f_{\max}) = 0$. The CT signals are filtered by well-tuned antialias filters (cutoff frequency below $f_s/2$)) before they are sampled. These should eliminate the signal power above half the sampling frequency $f_s/2$ to a user-specified level to keep the alias errors under control. The high frequency $(f > f_s/2)$ content of the measured signals is folded down in the frequency band of interest and acts as a disturbance. For this reason, it is strongly advised to always use antialias filters in the measurement set-up. Outside of digital control, the BL setup is the standard choice for DT measurements.

Measurement Setup—Notations

The general BL setup given in Figure 2 is considered the standard setup. The alternative simplified ZOH setup is obtained as a special case of the BL setup, by using the known DT generator sequence $r_d(k)$ in the memory of the generator as the measured input and removing the antialias filters.

The system is excited by the generator output signal $r(t)$ applied to the plant using the actuator. The generator signal is disturbed by generator noise $n_g(t)$, and the output of the system is disturbed by the process noise $n_p(t)$. The CT input and output signals are first low-pass filtered by the antialias filters. The measurement noise on the filtered input and output is, respectively, $m_u(t)$ and $m_y(t)$. These signals are sampled at rate $f_s = 1/T_s$. Eventually, the DT measurements $u(kT_s), y(kT_s), k = 1, \ldots, N$ will be used as the raw data from which the IR and FRF estimates will be obtained.

For notational simplicity, the difference between CT and DT signals will be no longer explicitly indicated. In the remainder of the article, the sampled signals $u(kT_s), y(kT_s)$ will be denoted as $u(k), y(k)$.

The DFT of the measurements $u(k), y(k), k = 1, \ldots, N$ is calculated using the fast Fourier transform (FFT) [34], and denoted as $U(k), Y(k)$. The frequency index k indicates the frequency kf_s/N. Because of the presence of the antialias filters, this setup is called the BL setup.

Noise Assumptions

The measured time-domain signals are

$$\begin{aligned} u(k) &= u_0(k) + n_u(k), \\ y(k) &= y_0(k) + n_y(k), \end{aligned} \tag{6}$$

where u_0, y_0 are the disturbance free signals and $n_u(k), n_y(k)$ model the combined disturbance contributions on, respectively, the input and output. The noise sequences n_u, n_y are assumed to be stationary sequences that can be mutually dependent (the input and output noise can be related). Without any loss of generality, they are assumed to be generated as filtered white noise, that is, $n_y = h_y * e_y$ where e_y is a DT white noise source. Although this simplified DT description of CT stochastic signals is very common in the system identification community, it is not obvious for this conceptual step. A profound theoretical foundation for this simplified representation is given in [16] and [35]

Assumption 1

In the time domain, the disturbing noise is $n = h * e$, where the white noise $e \sim N(0, \sigma_e^2)$. Different noise sources can be mutually correlated.

Remark 1

The Gaussian assumption is not needed for many of the results. The actual distribution becomes important to obtain quantified uncertainty bounds in the time domain for very small data sets. Asymptotically, the impact of the distributions on the uncertainty bounds disappears, and a Gaussian setting can be used.

Similar assumptions are made in the frequency domain

$$\begin{aligned} U(k) &= U_0(k) + N_U(k), \\ Y(k) &= Y_0(k) + N_Y(k), \end{aligned} \tag{7}$$

Design of Excitation Signals

Figure S1 highlights some important properties of excitation signals that are discussed in more detail below.

- *Random excitations*: Filtered random noise excitations (a) and (d) have a random amplitude spectrum for a given realization. At the dips of the amplitude spectrum, the frequency response function (FRF) estimate is most sensitive to disturbing noise (see the section "FRF Measurements Using Random Excitations" and Figure 7).
- *Periodic excitations*: Periodic signals give access to a general and detailed nonparametric noise analysis (see the section "Frequency Response Function Measurements Using Periodic Excitations") without any user interaction. The signals (c) and (f) have a deterministic amplitude spectrum that can be set by the user. Such signals are not prone to dips in the amplitude spectrum that vary from one realization to the other. This results in a better guaranteed signal-to-noise ratio (SNR), even for a single realization. Leakage errors need to be avoided/removed by processing an integer number of periods (see the section "Frequency Response Function Measurements Using Periodic Excitations") or by using the advanced algorithms of the section "Improved Frequency Response Function Measurements Using Local Parametric Methods."
- *Deterministic amplitude spectrum*: All the signals that are shown in Figure S1 have a random nature, even if the amplitude spectrum can be deterministic. This makes these signals well suited to be used as an excitation signal for FRF measurements in the presence of nonlinear distortions. The random nature is smoothing the impact of the nonlinear distortions so that the FRF still represents the averaged linearized behavior of the system [30].
- *Special designed excitation signals*: A number of special signals are discussed next: filtered noise, maximum length binary sequence (MLBS), pseudorandom binary sequence (PRBS), swept sine, and multisines. An extensive discussion of excitation signals can be found in [16], [18], [25], and [29].

MAXIMUM LENGTH BINARY SEQUENCE

The MLBS belongs to the class of PRBSs, which are deterministic, periodic sequences of length N that switch between one level (such as +1) and another (such as –1). The switches can occur only on an equidistant grid at multiples of the clock period T_c, and they are chosen such that the autocorrelation is as spiky as possible to mimic a Dirac impulse (see Figure S4) [25], [26]. It is not possible to find a binary sequence with these properties for every arbitrary length N. For example, MLBSs exist only for length $2^n - 1$. In [27], an overview is given of binary and near-binary sequences that give the user a large choice of sequence lengths beyond that of the MLBS. Publicly available software to generate these signals is discussed in [28].

The continuous-time (CT) sequence is obtained using a zero-order hold (ZOH)-reconstruction of the discrete-time sequence, as shown in Figure S3(a) for an MLBS. The amplitude spectrum of the discrete sequence is a constant (except for the dc value). The spectrum of the CT sequence rolls off with

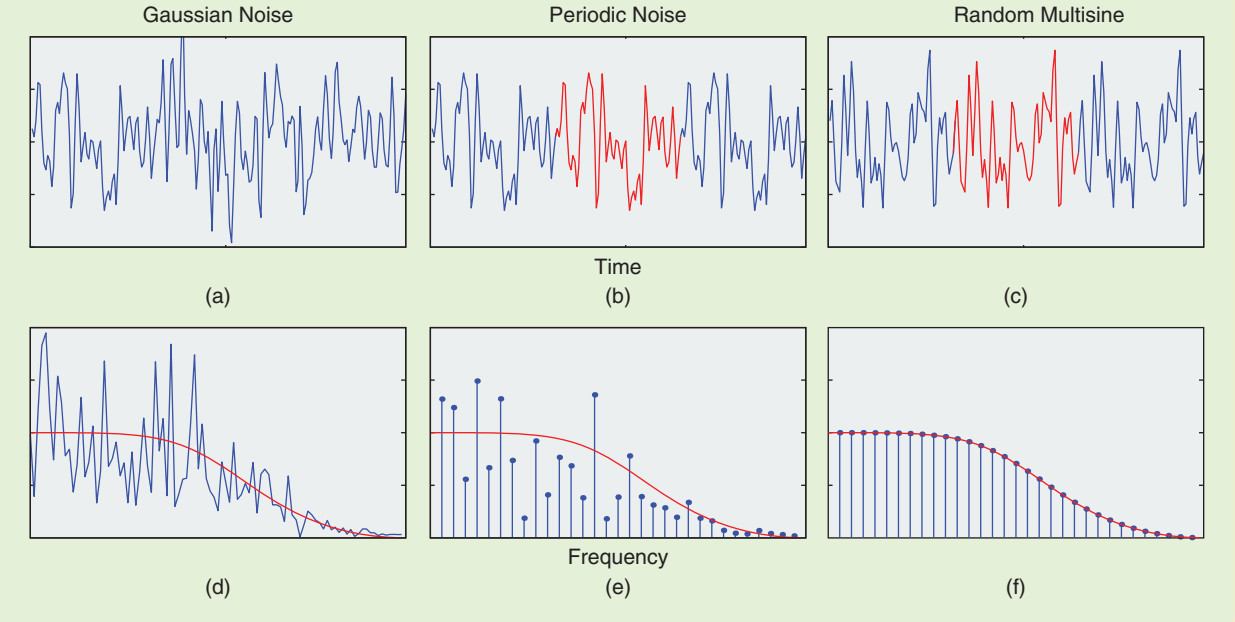


FIGURE S1 The characteristics of excitation signals are shown in the time and the frequency domain for (a) and (d) Gaussian noise, (b) and (e) periodically repeated Gaussian noise, and (c) and (f) random-phase multisine. In the frequency domain, the amplitude spectrum of the actual realization (blue) and the power spectrum (red) are shown.

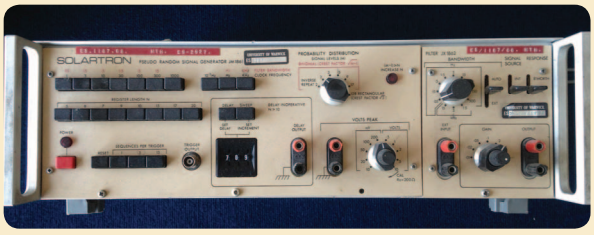


FIGURE S2 A pseudorandom binary signal generator from the 1960s used as an example in laboratories at the School of Engineering, University of Warwick. Dedicated generators to generate maximum length binary sequence and related signals like the inverse repeat binary sequence were built and commercialized. Besides the original signal $u(t)$, a copy with a user adjustable delay $u(t-\tau)$ was also generated. This allowed the correlation $y(t)u(t-\tau)$ to be measured using analog correlators, making a direct measurement of the impulse response $g(\tau)$ possible.

the $ZOH(f) = \sin(\pi f)/(\pi f)$ characteristic [see Figure S3(d)]. This restricts the useful frequency band, because the $ZOH(f)$ has zeros at the multiples of the clock frequency $f_c = 1/T_c$. Moreover, the signal level drops with the frequency, which results in a decreasing SNR for higher frequencies. In the 1960s, dedicated generators were used to generate these signals (see Figures S2 and S13).

User Guidelines

- *Choice of the clock frequency*: Select $f_c = 2.5\, f_{\max}$ to obtain a sufficiently flat amplitude spectrum in the frequency band of interest [16].
- *Signal length*: Select the length N of the sequence, such that $f_0 = 1/N$ meets the frequency resolution requirement.
- *Fast Fourier transform (FFT) analysis*: Do not modify the signal by padding the signal with zeros to obtain a period length that is a power of two. Instead, generalized FFT algorithms should be used that can handle arbitrary signal lengths. These are widely available in commonly used routines. Zero padding will destroy the good spectral properties of the MLBS.
- *Eliminate even nonlinearities*: The major drawback of MLBSs or more general PRBSs is the sensitivity to nonlinear distortions. These create large spikes in the impulse response estimates [26], [102]. A first possibility to remove these spikes is to calculate the median over multiple randomized realizations. Alternatively, an inverse repeated sequence $[u, -u]$ can be generated. All even nonlinear distortions are eliminated by construction at a cost of a reduction of the frequency resolution with a factor two.
- *Ternary signals*: An increased robustness against nonlinear distortions is obtained using well-designed ternary signals [103]. These signals excite only a set of selected odd frequencies. The remaining unexcited odd frequencies, and the even frequencies, can be used to measure the level of the nonlinear distortions [30].

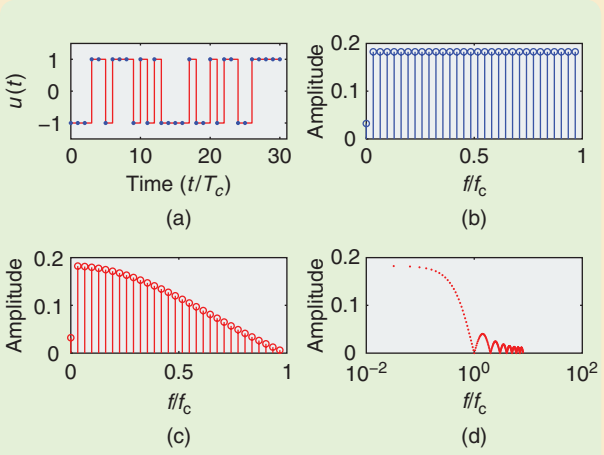


FIGURE S3 An example of a maximum length binary sequence (MLBS) with a length $N = 2^5 - 1 = 31$. (a) Discrete-time (DT, blue dots) and continuous-time (CT, red line) MLBS with a clock frequency $f_c = 1/T_c = 1$. The CT sequence is obtained by a zero-order hold reconstruction. (b) The spectrum of the DT sequence, with $f_s = f_c$. The spectrum of the DT sequence is perfectly flat (except for the dc value). This corresponds to the observation that the discrete correlation function consists within a dc offset of a perfect Dirac, as shown in Figure S4. The spectral resolution is f_c/N, which increases with the length of the sequence. The amplitude will drop as $1/\sqrt{N}$ because the power is equally spread over N frequencies. Parts (c) and (d) show the amplitude spectrum of the CT MLBS. The spectrum is proportional to $\sin(\pi x)/(\pi x)$ with $x = f/f_c$. The amplitude spectrum has a zero at the multiples of f_c.

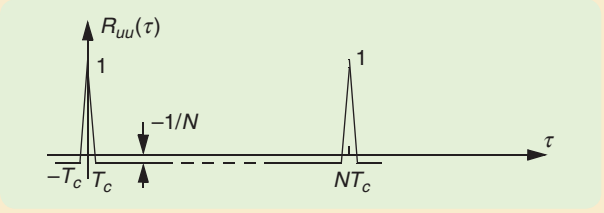


FIGURE S4 The autocorrelation of a continuous-time maximum length binary sequence with levels ±1 and length $N = 2^n - 1, n \in \mathbb{N}$. The autocorrelation is periodic with period length N. Note the offset of $1/N$. The width of the impulse is set by the clock period T_c. For the discrete-time sequence, the autocorrelation exists only for $\tau = kT_c$. In that case, it becomes a perfect Dirac function with an offset of $-1/N$.

GENERAL-PURPOSE EXCITATIONS

Figure S5 shows other general purpose excitation signals besides the MLBS. The full details about these signals can be found in [16], [24], and [29]. The signals were designed to cover the frequency band [1, 50] Hz in a measurement window of 1 s.

Swept Sine or Chirp Excitation

This is a sine excitation with an instantaneous frequency that is periodically linearly varying between f_1 and f_2

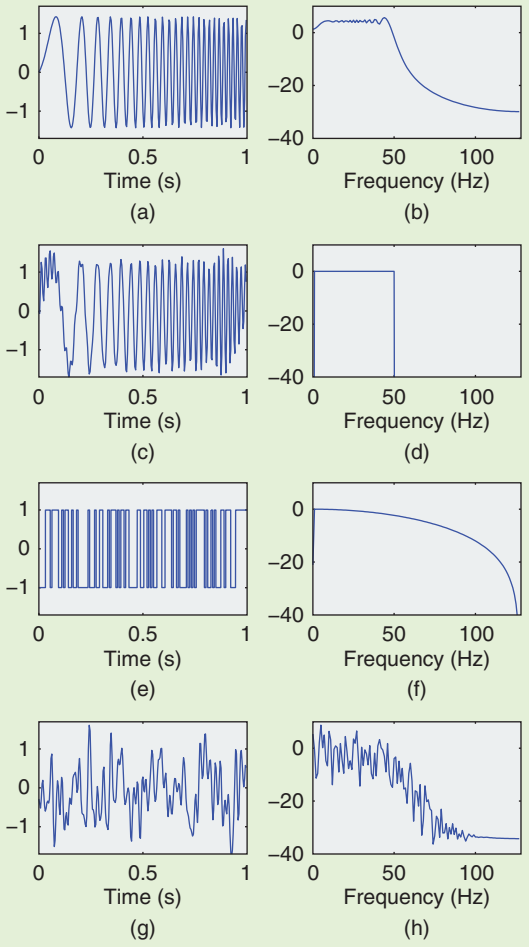


FIGURE S5 A comparison of general-purpose excitation signals in the time and frequency domains. The aim is to excite a frequency band between 1 and 50 Hz using signals with a length of 256 samples and a generator clock frequency of 256 Hz. (a) Swept sine, (c) Schroeder multisine, (e) maximum length binary sequence (MLBS), and (g) filtered white noise. For the MLBS, a clock frequency of 127 Hz was used.

$$u(t) = A\sin((at+b)t), 0 \le t < T_0, \tag{S3}$$

with $T_0 = 1/f_0$ the period, $a = \pi(k_2 - k_1)f_0^2, b = 2\pi k_1 f_0$, $f_1 = k_1 f_0$, $f_2 = k_2 f_0$. We note that, in Figure S5(b), some of the power is outside the frequency band of interest. Inside the frequency band, a ripple of a few decibels is present.

Multisine

A multisine is the sum of harmonically related sines

$$u(t) = \sum_{k=1}^{F} A_k \cos(2\pi k f_0 t + \phi_k). \tag{S4}$$

The user can freely choose the amplitudes A_k and the frequency resolution f_0. The choice of the phases will set the nature of the signal. Random phases between $[0, 2\pi)$ will create a Gaussian-like behavior. The phases can also be optimized to minimize the peak value of the signal [45]. In Figure S5(c) and (d), the amplitudes were put equal to a constant, and the phases were $\phi_k = -k(k-1)\pi/F$. This is a Schroeder multisine [46] that mimics a swept sine, with a perfectly flat amplitude spectrum in the frequency band of interest and no power outside this band. This comes at a cost of an increased peak value in the frequency domain. The latter can be reduced by phase optimization algorithms [45]. Multisines are the most flexible class of periodic excitation signals [16], [24].

Filtered White Noise

As discussed in the section "Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods," the spectrum of filtered white noise in Figure S5(h) is also a random variable. At some frequencies, almost no power will be present, resulting in a low SNR. For that reason, it is strongly advised to average the FRF-measurements over a number of realizations (subrecords) to obtain good results. It can also be observed that the time-domain signal in Figure S5(g) has large peak values. This is typical for Gaussian noise excitations signals.

where U_0, Y_0 are the disturbance free signals and $N_U(k), N_Y(k)$ model the combined disturbance contributions on respectively the input and output at frequency k.

Assumption 2

In the frequency domain, the disturbing noises $N_U(k), N_Y(k)$, for all frequencies k, are complex, circular, and normally distributed [16] with the following properties:

$$\begin{aligned} E\{N_U(k)\} = 0, E\{N_Y(k)\} = 0, \\ E\{|N_U(k)|^2\} = \sigma_U^2(k), E\{|N_Y(k)|^2\} = \sigma_Y^2(k), \\ E\{N_Y(k)\bar{N}_U(k)\} = \sigma_{YU}^2(k), E\{N_Y(k)N_U(k)\} = 0. \end{aligned} \tag{8}$$

In the last expression, $\bar{x}$ denotes the complex conjugate of x.

Assumption 3

The noise $N_U(k), N_Y(k)$ is independent of $U(l), Y(l)$, for all $k \neq l$.

The role of the generator noise in Figure 2 differs depending on the selected processing approach, and it also affects the definition of u_0, y_0. In the *periodic framework* (where the excitation is assumed to be periodic), $u_0(t), y_0(t)$ are the signals that are solely due to the reference excitation $r(t)$. Using the notation of Figure 2, $u_1(t) = u_0(t)$ and $y_1(t) = y_0(t)$ if all noise sources n_g, n_p, m_u, m_y are equal to zero. The generator

noise will act as a nonperiodic disturbance that does not affect the uncertainty of the estimates (see the section "Special Case: Generator Noise Only"). In the *nonperiodic framework* (using, for example, random excitations), $u_0(t), y_0(t)$ are the signals that are solely due to the reference excitation $r(t)$ and the generator noise $n_g(t)$. Here, $u_1(t) = u_0(t)$, and $y_1(t) = y_0(t)$ if the noise sources n_p, m_u, m_y are equal to zero. The generator noise $n_g(t)$ can be considered to be a part of the excitation that is out of user control.

User Guidelines

The experimental setup has a significant impact on the final quality. Making small modifications can simplify and improve the processing of the data and the quality of the results. For this reason, the user must pay attention to the following aspects:

- *Antialias filters*: Verify if antialias filters are present, and make a proper selection of the cut-off frequency.
- *Synchronization*: Make sure that the data acquisition and data generation channels are well synchronized. Lack of synchronization can jeopardize the quality of the data. Nonparametric, periodic postprocessing heavily relies on a perfect synchronization. If the clocks of the data acquisition and the generator are not well synchronized, it is still possible to resample the signals using advanced signal processing algorithms that estimate the clock mismatch and the corrected discrete Fourier spectra [36].
- *Preprocessing*: Check what manipulations are applied to the data: drift/trend removal, missing data handling, removal of outliers, and prefiltering.
- *Reference signal*: Store the reference signal together with the measured data. It becomes much easier to avoid systematic errors if the reference signal is available, especially for closed-loop measurements.
- *Signal assumption*: Make sure that the measurement setup is in agreement with the ZOH (antialias filters switched off) or BL (antialias filters switched on) intersample assumption.

FREQUENCY RESPONSE FUNCTION MEASUREMENTS USING PERIODIC EXCITATIONS

This section analyzes the FRF measurement in full detail using periodic excitations. Initially, the FRF was measured frequency per frequency using a sine excitation [37]. By moving to periodic excitations that excite multiple frequencies at once, it was possible to significantly reduce the measurement time [38]. Such periodic signals are available in all commercial dynamic signal analyzers. Their use was strongly stimulated by the introduction of the Fourier transform in the electrical engineering field [39]. The development of efficient numerical procedures to calculate the Fourier transform was the start of a new era [40]–[43]. Extending the study to include random excitations will be discussed in the next section, leading to the classic FRF measurement methods that have dominated the field since the 1960s.

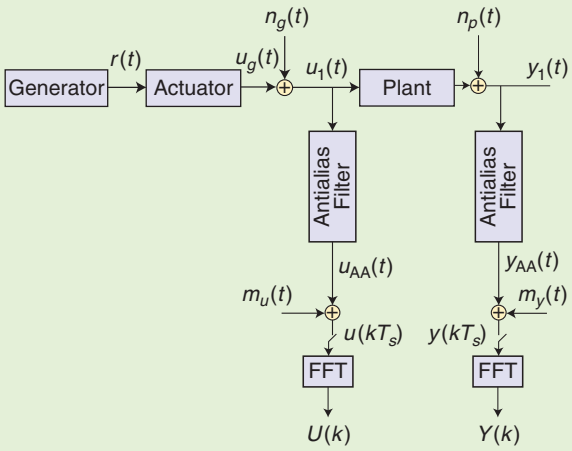


FIGURE 2 Measurement setup and notations. The generator signal is disturbed by generator noise $n_g(t)$, and the output of the system is disturbed by the process noise $n_p(t)$. The measured input and output signals are first low-pass filtered by the antialias filters. The measurement noise on the input and output is, respectively, $m_u(t)$ and $m_y(t)$. These signals are sampled at rate $f_s = 1/T_s$. The discrete Fourier transform of the measurements $u(kT_s), y(kT_s), k = 1, \ldots, N$ is calculated using the fast Fourier transform (FFT) algorithm, and it is denoted as $U(k), Y(k)$. The frequency index k indicates the frequency kf_s/N.

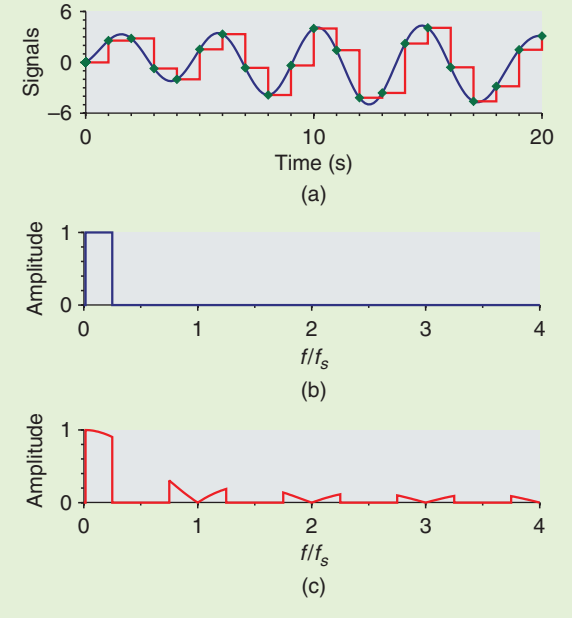


FIGURE 3 A band-limited (BL) and zero-order hold (ZOH) reconstruction. (a) The time signals, with green diamonds representing the samples, the blue line representing the BL reconstruction, and the red line representing ZOH reconstruction; (b) the amplitude spectrum of the BL limited reconstruction; and (c) the amplitude spectrum of the ZOH reconstruction. Observe the high-frequency repetitions around the multiples of f_s.

In both the time- and the frequency-domain approach, the choice of the excitation signal has an impact on the methods to be used.

Note that the measured FRF is a random variable (see "Random Nature of Frequency Response Function Measurements"), the stochastic behavior of which will be characterized by calculating the mean value and the variance (see "Characterizing the Stochastic Properties of the Frequency Response Function Estimates").

Consider the SISO system in Figure 1. The frequency-domain analysis begins with the measured input and output DFT spectra (7). To simplify the expressions, the frequency index k will be omitted. Only when multiple frequencies are combined or when it is not clear from the context what frequency is used will the frequency index will be added.

Stochastic Analysis of Periodic Excitations

The input $u(t)$ and output $y(t)$ are measured over multiple periods and broken into subrecords of one period each. For example, $u^{[l]}, l = 1, \ldots, P$, as shown in Figure 4. In the next step, the mean value and the (co)variance are calculated as a function of the frequency by analyzing the variations of the periodic input and output signals over the measurements of the repeated periods. While the disturbing noise n_u, n_y varies from one period to another, the noiseless periodic signals u_0, y_0 do not. This results eventually in the following simple procedure. For each subrecord $u^{[l]}, y^{[l]}$ (corresponding to one period), the DFT $U^{[l]}, Y^{[l]}$ is calculated using the FFT algorithm. If an integer number of periods is measured under steady-state conditions (no transients present), there will be no leakage in the results. The sample mean and noise (co)variances at frequency k are then

$$\hat{U}(k) = \frac{1}{P}\sum_{l=1}^{P} U^{[l]}(k) \quad \hat{Y}(k) = \frac{1}{P}\sum_{l=1}^{P} Y^{[l]}(k), \tag{9}$$

and

$$\begin{aligned}
\hat{\sigma}^2_U(k) &= \frac{1}{P-1}\sum_{l=1}^{P} \left| U^{[l]}(k) - \hat{U}(k) \right|^2, \\
\hat{\sigma}^2_Y(k) &= \frac{1}{P-1}\sum_{l=1}^{P} \left| Y^{[l]}(k) - \hat{Y}(k) \right|^2, \\
\hat{\sigma}^2_{YU}(k) &= \frac{1}{P-1}\sum_{l=1}^{P} \left(Y(k) - \hat{Y}(k) \right)\left(\bar{U}(k) - \bar{\hat{U}}(k) \right).
\end{aligned} \tag{10}$$

Note that the variance of the estimated mean values $\hat{U}(k), \hat{Y}(k)$ is given, respectively, by $\sigma^2_U(k)/P, \sigma^2_Y(k)/P$. Combining this

Random Nature of Frequency Response Function Measurements

Frequency response function (FRF) measurements $\hat{G}(k)$ (often called FRF estimates) that are obtained from finite-length measurements differ from the true value $G_0(k)$ by an error term $N_G(k)$, with $\hat{G}(k) = G_0(k) + N_G(k)$ because the estimation process is disturbed by many error sources (see Figure 2). The measured input and output are prone to measurement noise $m_u(t), m_y(t)$; unknown inputs can act on the system to be modeled, which leads to process noise $n_p(t)$. Finite measurement length effects also disturb the estimate $\hat{G}(k)$ because the past inputs (before the start of the experiment) and the future outputs (after finishing the measurements) are not always properly included in the calculations.

A typical FRF measurement is shown in Figure S6. It can be observed that the errors vary rapidly from one frequency to another. Repeating the experiment leads to a new FRF measurement $\hat{G}(k)$ that differs from the previous one because the input changed (for random excitations). The measurement and process noise also varies from one experiment to the other.

The error term $N_G(k)$ is modeled as a random variable that can be studied using statistical tools. The process of generating a new noise sequence each time is called a "realization" of the noise.

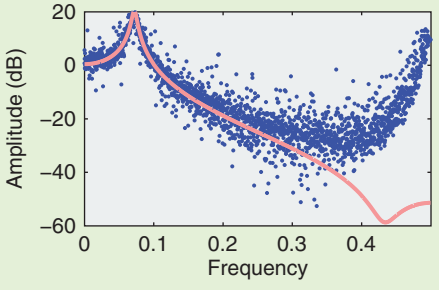


FIGURE S6 Simulation results on a second-order, discrete-time system. Frequency response function (FRF) measurement starting from a finite-length record of undisturbed data uses a filtered random noise excitation with a bandwidth of $0.4\,f_s$. The pink line is true FRF G, and the blue dots are the FRF measurement. Even in the absence of disturbing noise, the measurements look very noisy due to the leakage effect. In this case, the response of a second-order system with a time constant of eight samples was simulated in $N = 4096$ samples.

information in one figure results in a full nonparametric analysis of the SNR of the measurements. No interaction with the user is needed during the processing. This makes the method well suited for implementation in standard measurement procedures. The FRF estimate is then $\hat{G} = \hat{Y}/\hat{U}$, the properties of which are studied in the section "Smoothing Frequency Response Function Measurements for Periodic Excitations."

Characterizing the Stochastic Properties of the Frequency Response Function Estimates

A random variable is fully characterized in amplitude by its probability density function (pdf) [3], [55]. It is often too difficult to obtain the pdf of $N_G(k)$, and for that reason the pdf is replaced by partial information: its mean value $\mu(k) = E\{N_G(k)\}$ and its variance $\sigma^2 = E\{|N_G(k) - \mu(k)|^2\}$, where $E\{x\}$ is the expected value of x [55].

If $\mu(k) \neq 0$, the frequency response function measurement $\hat{G}(k)$ is prone to a systematic error that cannot be removed by averaging the result over multiple measurements. Such an error is called a bias. It is often desired to make the bias as small as possible because it cannot be (easily) removed in the remaining data processing. This is further discussed in "Bias and Variance Tradeoff of Estimators."

The variability of $\hat{G}(k)$ around its expected value is characterized by the variance σ^2. The larger the variance, the wider the spread around the expected value. For a Gaussian pdf, the interval $[-1.96\sigma, +1.96\sigma]$ corresponds to the 95% confidence interval.

In practice, the confidence of the measurement $\hat{G}(k)$ is often given by drawing either the $[-1.96\sigma, +1.96\sigma]$ interval around the measured value or by plotting $\sigma(k)$, as was done in Figure 6. These intervals do not account for the presence of a bias (systematic errors).

Bias and Variance Tradeoff of Estimators

An error e can always be written as the sum of its mean value $b = E\{e\}$ (called the bias), and the remainder is $v = e - b$ with variance σ^2, such that $e = b + v$. The total mean square error is

$$e_{MS} = b^2 + \sigma^2. \tag{S5}$$

Depending on the preference, either the bias b or the mean square error e_{MS} should be as small as possible. It is always possible to scale an unbiased estimator (no bias present) toward zero such that e_{MS} drops. This is illustrated in Figure S7 on the following simple scalar example. Assume that $\hat{\theta}$ is an unbiased estimate of the true parameter $\theta_0 = 1$, with variance $\sigma^2 = 1$. Consider next the scaled estimator $\tilde{\theta} = \lambda\hat{\theta}$. The bias of $\tilde{\theta}$ is $b = (1 - \lambda)$, and the variance $\tilde{\theta}$ is $\sigma_{\tilde{\theta}}^2 = \lambda^2$. The mean square error becomes

$$e_{MS} = (1 - \lambda)^2 + \lambda^2. \tag{S6}$$

This error, plotted in Figure S7 as a function of the scaling factor λ, shows a minimum at $\lambda = 0.5$.

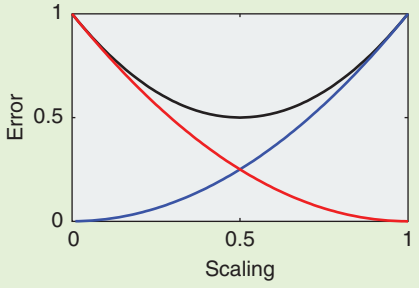


FIGURE S7 The evolution of the total mean square error as a function of the squared bias and variance error. The bias-variance error tradeoff is tuned by scaling an unbiased estimator with a scaling factor between zero and one. Black line: mean square error, red line: bias error, and blue line: variance error.

USER GUIDELINE BIAS AND VARIANCE TRADEOFF FOR IMPULSE RESPONSE AND FREQUENCY RESPONSE FUNCTION MEASUREMENT

Depending on the need of the user, it is preferable to tune the results to have either a low bias or variance.

- If the nonparametric impulse response or frequency response function estimates will be used as the input of a parametric modeling step, it is most important that no bias errors are present because these cannot be removed anymore in later postprocessing. The parametric estimation step is considered an advanced smoothing algorithm that reduces the noise without introducing a bias error if it is well designed [16], [21], [22].
- To generate initial estimates for a parametric estimation step, a bias can be tolerated [75]. The final estimation should be completed on the original, not smoothed, bias-free data.
- If the nonparametric results will be used, it is important that the combined bias/variance error is as small as possible, so that the mean square error is a better measure of the quality.

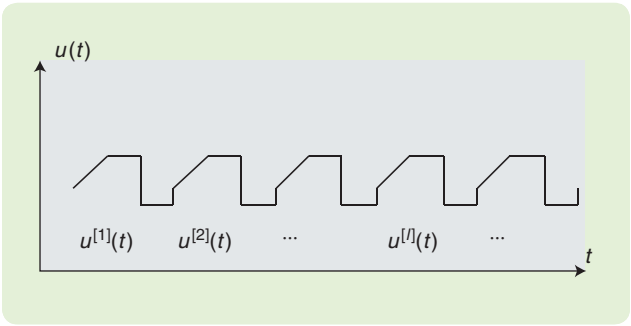


FIGURE 4 Calculating the sample mean and sample variance of a periodic signal starting from multiple measured periods. The input $u(t)$ and output $y(t)$ are measured over multiple periods and broken into subrecords of one period each, for example $u^{[l]}, l = 1, \ldots, P$. The fast Fourier transform is applied to each subrecord. In the next step, the mean value and the (co)variance are calculated as a function of the frequency.

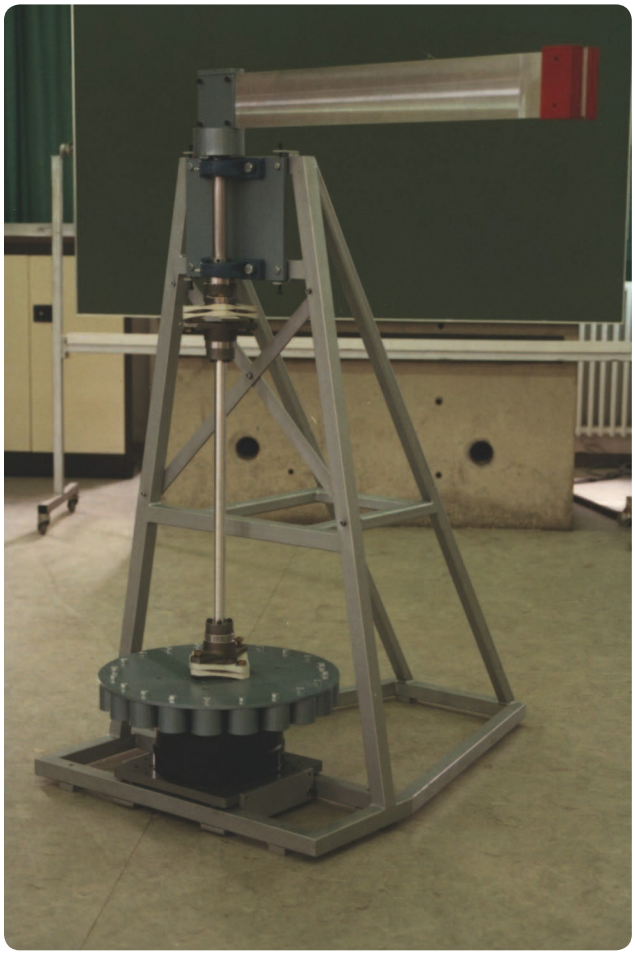


FIGURE 5 The one degree-of-freedom flexible robot arm of the KU Leuven PMA Department of Mechanical Engineering. The red mass at the tip of the flexible arm is driven by an electric motor. The input signal is the motor current, and the output signal is the acceleration measured at the tip of the robot. A periodic excitation signal is used, the period length is $N = 4096$ with $f_s = 500$ Hz, and $P = 10$ periods are measured in steady-state conditions. Only the odd frequencies $[1, 3, \ldots, 201]\, f_0$, with $f_0 = f_s/N = 0.1221$ Hz are excited.

Example: The Flexible Robot Arm

As an example, experimental data measured under the BL assumption on a one degree-of-freedom flexible robot arm (see Figure 5 and [44] for a detailed description) are processed using (9) and (10). A periodic multisine excitation [45], [46] is used (see also "Design of Excitation Signals"). In Figure 6, the sample mean $\hat{U}, \hat{Y}$ and standard deviation $\hat{\sigma}_U$, $\hat{\sigma}_Y$ are plotted in (a) and (b), respectively. The uncertainty on the mean values will be 10 dB below the actual shown noise levels because ten periods are averaged (see "Impact of Averaging on the Variance of the Frequency Response Function Estimate"). These results show that, during the measurement process, an excellent impression of the quality of the measurement is obtained

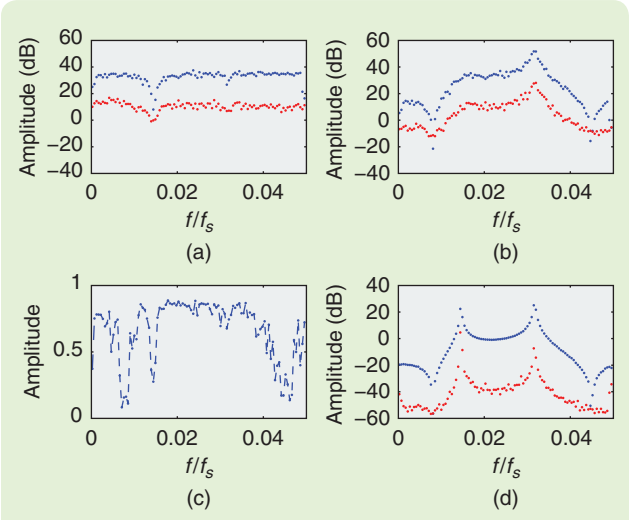


FIGURE 6 Processing the periodic measurements on the robot arm in Figure 5. (a) The discrete Fourier transform (DFT) spectrum U of the current, (b) the DFT spectrum Y of the acceleration, (c) the correlation ρ (11), and (d) the frequency response function obtained by (12) $\hat{G} = \hat{Y}/\hat{U}$ (see the section "Smoothing Frequency Response Function Measurements for Periodic Excitations"). Blue shows the amplitude, and red shows the standard deviation.

Impact of Averaging on the Variance of the Frequency Response Function Estimate

Averaging an estimate over P independent realizations (that is, one measurement has nothing to do with the other) results in a reduction of the variance σ^2 by a factor P, or a reduction of the standard deviation by $\sqrt{P}$

$$\sigma_{\hat{G}_{\text{aver}}} = \sigma_{\hat{G}}/\sqrt{P}. \tag{S7}$$

Averaging can be done by repeating the experiment P times. This is a very simple and attractive solution to reduce the standard deviation, but it comes with a price. The measurement time grows proportional to P, while the uncertainty drops only by $\sqrt{P}$. Reducing the standard deviation by a factor of ten increases the measurement time by a factor of 100.

without any user interaction. In this case, the SNR is approximately 20 dB before averaging, except for the input of $f/f_s = 0.015$ corresponding to the first resonance frequency of the robot arm. The drop in the SNR is due to the impedance mismatch between the motor and the structure, creating a power drop at the input and resulting in a low SNR at the input (about 0 dB before averaging). This is also why it is often very difficult to make a precise damping measurement of lowly damped mechanical structures.

The correlation ρ measures the linear relation between two noise sources

$$\rho = \frac{\sigma^2_{YU}}{\sigma_U \sigma_Y}, \tag{11}$$

and is shown in Figure 6(c). In this case, the high correlation (values close to one) indicates that the noise on the force and accelerator measurements are highly correlated. This can be due to either generator noise (which is not in this setup) or the presence of a dominant noise source in a closed loop. The latter is the case in this setup. The output of the motor is strongly loaded by the input of the mechanical structure, leading to an internal feedback mechanism. The measured FRF shown in Figure 6(d) is obtained using the methods discussed in the next section. The correlation is much lower at the first resonance and antiresonance because, at those frequencies, either the input or the output drops to very low values, as shown in Figure 6(a) and (b), and the measurement noise will then become dominant. The latter is not correlated with the noise in the loop.

User Guidelines

- *Periodic inputs*: Use periodic inputs whenever it is possible.
- *Number of repeated measurements*: The (co)variance estimates improve with the number of repeated periods. However, in practice a small number of repetitions ($P \geq 4$) will suffice. If the variances are used in a parametric estimate, then $P \geq 4$ is enough to guarantee consistency of the estimates. $P \geq 7$ guarantees the existence of the covariance matrix of the estimated parameters and their limiting distribution [16], [47].
- *Overlapping subrecords*: These results can be generalized to measurements with only two repeated periods using overlapping subrecords [48].
- *Covariance information*: It is strongly advised to calculate the covariances, because these provide valuable information about the presence of generator noise or feedback loops.

Smoothing Frequency Response Function Measurements for Periodic Excitations

Smoothing: Average over the Periods

The simplest FRF estimate $\hat{G}$ is the empirical transfer function estimate (ETFE) [21], starting from the sample mean $\hat{U}, \hat{Y}$ (9)

$$\hat{G} = \frac{\hat{Y}}{\hat{U}} = \frac{Y_0 + N_{\hat{Y}}}{U_0 + N_{\hat{U}}} = G_0 \frac{1 + N_{\hat{Y}}/Y_0}{1 + N_{\hat{U}}/U_0}, \tag{12}$$

with $G_0 = Y_0/U_0$. This is a "local" nonparametric method; the estimates $\hat{G}(k)$ at frequency k make no use of information at other frequencies $k \neq l$. All the estimates are retrieved by repeating this calculation at all frequencies of interest.

Bias Expression

Creating the Taylor series expansion of (12) to degree two and using the noise Assumption 3, the bias is

$$\begin{aligned} b &= E\{\hat{G}\} - G_0, \\ &= -G_0 \exp\left(-|U_0|^2/\sigma^2_{\hat{U}}\right)\left(1 - \rho \frac{U_0/\sigma_{\hat{U}}}{Y_0/\sigma_{\hat{Y}}}\right), \\ &= -G_0 \exp\left(-P|U_0|^2/\sigma^2_U\right)\left(1 - \rho \frac{U_0/\sigma_U}{Y_0/\sigma_Y}\right). \end{aligned} \tag{13}$$

See [49] for $\rho = 0$ and [50] for the correlated noise situation. Even for SNR values as low as 10 dB, the relative bias $|b/G_0|$ is still below 10^{-4}. From (13) it follows that the bias on the FRF decreases exponentially with the number of averaged periods P. The order of the operations is important. First the spectra should be averaged, and the division can then be calculated. If the order is reversed, there will be no bias reduction. This leads to a general rule of thumb: the data should first be averaged before nonlinear operations like division or multiplication are applied. Failing to do so makes the method more prone to systematic errors. For nonsynchronized measurements, it becomes difficult to apply an average before the division. In that case, nonlinear averaging methods can be used to reduce the bias [49], [51] at a (small) loss in efficiency. For this reason, synchronized measurements are still preferred whenever possible.

Practical Variance Expression: Study of the Distribution of $\hat{G}$

In [52] and [53], it is shown that the variance of (12) does not exist, due to the presence of outliers when the denominator closely approaches zero. Nevertheless, it is still possible to provide exact uncertainty intervals on the FRF measurement, starting from the probability density distribution of $\hat{G}$. The exact expression is known and studied in detail in [52] and [54]. A detailed discussion of these results is beyond the scope of this article. Instead, the practical conclusions of the full analysis are discussed.

Distribution

Under the Gaussian noise Assumption 2 and for a high SNR of the input signal (> 40 dB), the FRF $\hat{G}$ is an approximate, complex Gaussian distribution because, in (12), the division is $1/(1 + N_Y/\hat{U}) \approx 1 - N_Y/\hat{U}$. For a lower input SNR, the linear approximation no longer holds and the higher-order terms in the approximation will affect the distribution.

Uncertainty Bounds

It is still possible to calculate a practical variance expression of $\hat{G}$ that can be used to generate uncertainty bounds assuming a complex Gaussian distribution for $\hat{G}$. The higher the input SNR, the tighter the uncertainty interval is. For example, a tight 50% uncertainty interval is obtained for an SNR that is larger than 10–15 dB, and a tight 95% uncertainty bound requires an SNR that is larger than 20 dB. The variance expression is obtained from the first-order Taylor series expansion of (12), under Assumption 3 [16]

$$\sigma^2_{\hat{G}} = |G_0|^2\left(\frac{\sigma^2_{\hat{Y}}}{|Y_0^2|} + \frac{\sigma^2_{\hat{U}}}{|U_0^2|} - 2\,\mathrm{Re}\left(\frac{\sigma^2_{\hat{Y}\hat{U}}}{Y_0\bar{U}_0}\right)\right),$$
$$= \frac{1}{P}|G_0|^2\left(\frac{\sigma^2_{\hat{Y}}}{|Y_0^2|} + \frac{\sigma^2_{\hat{U}}}{|U_0^2|} - 2\,\mathrm{Re}\left(\frac{\sigma^2_{\hat{Y}\hat{U}}}{Y_0\bar{U}_0}\right)\right). \quad (14)$$

The variance is inversely proportional to the squared SNR of the measurements, and it drops in $1/P$. In practice, the variance is estimated by replacing U_0, Y_0 by $\hat{U}, \hat{Y}$ (9) and the (co)variances by the sample (co)variances (10).

This result will guide the excitation design in "Design of Excitation Signals." Good excitation signals should maximize the SNR at all frequencies of interest. Doubling the SNR allows the measurement time to be reduced by a factor of four, so a good excitation design (increasing $|U_0|$) and a good measurement setup design (decreasing the variances) reduce the measurement time significantly.

Special Case: Generator Noise Only

If only generator noise n_g is present in Figure 2, $\sigma^2_{\hat{U}} = \sigma^2_{N_g}$, $\sigma^2_{\hat{Y}} = |G|^2\sigma^2_{N_g}$, and $\sigma^2_{\hat{Y}\hat{U}} = G\sigma^2_{N_g}$. Substituting these results in (14) shows that $\sigma^2_{\hat{G}} = 0$. In the periodic framework, the generator noise does not contribute to the variance of the FRF measurement. The same conclusion holds true in the general excitation framework, where the generator noise adds directly to the measured input signal and acts as an input that is not under user control.

Example: The Flexible Robot Arm—FRF Measurement

Both (12) and (14) are applied to the flexible robot arm data as previously discussed. The FRF estimate is calculated using the averaged input and output data, and the variance is calculated using the sample variances. The results are shown in Figure 6(d). A smooth FRF estimate is obtained, but it is also clearly visible that the uncertainty around the resonance (of the first mode) is very high. This is due to the poor SNR at the input in that frequency band, indicating once more that low damping values are very difficult to measure with high precision.

User Guidelines

- Use periodic excitations whenever it is possible. Check the synchronization between the generator and the data acquisition channels. Use advanced signal processing methods to remove synchronization errors [36].
- *Nonparametric noise analysis*: Use nonparametric noise analysis to complete an initial quality check of the measurements. Identify the dominant noise source (input or output noise). This information might be useful to refine the experiment design or improve the measurement setup.
- *Mutual correlation*: Check the cross-correlation between the input and output noise, which indicates if one or two independent noise sources are present in the setup. It can also be an indication of strong closed-loop effects, which should be handled with care (see "Measuring the Frequency Response Function Under Closed-Loop Conditions").
- *Steady-state conditions*: Check if the measurements are completed under steady-state conditions. The initial transient is estimated by subtracting the last period from the first period. If transients are present, more advanced methods should be used (see the section "Improved Frequency Response Function Measurements Using Local Parametric Methods").
- *Averaging*: Average the input and output DFT spectra before making the division, to reduce bias effects on the FRF estimate, as in (12).

FREQUENCY RESPONSE FUNCTION MEASUREMENTS USING RANDOM EXCITATIONS

This section repeats the discussion of the FRF measurement for random excitations. The first step is to introduce the basic problems, which are 1) why noiseless data can result in the poor measurements shown in Figure 7 and 2) how these errors can be removed using smoothing methods. Further discussions consider the variance and bias tradeoffs that are made when smoothing is applied. The classical methods that were developed during the 1960s to solve these are discussed in the section "Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods." Recent insights into the structured nature of leakage errors (see "Models for Dynamic Systems: Finite Length Effects") will be used to provide a deeper understanding of these well-established methods.

Smoothing Frequency Response Function Measurements for Random Excitations Using the Empirical Transfer Function Estimate

The poor results shown in Figure S6 were retrieved by applying the ETFE (12) with $P = 1$ (the data are not split into subrecords) on simulations obtained with a random noise excitation. No smoothing was applied before the division. Figure 7 plots the same results as shown in Figure S6, with the amplitude $|U|$ of the DFT of the actual realization of the input that was used for this measurement. The latter is also a random variable, having dips at some frequencies [4]. It is mainly at those frequencies that $\hat{G}$ becomes very prone to disturbances, resulting in the very scattered nature on the left side of the figure. In this example, the

Measuring the Frequency Response Function Under Closed-Loop Conditions

Measuring the frequency response function (FRF) of a system under closed-loop conditions (Figure S8) requires special precautions. Depending on the signal-to-noise (SNR) of the measurements, the resulting FRF is that of the feedforward pathway (or the inverse FRF of the feedback pathway) or a combination of both results. The FRF estimate starting from the cross and auto spectrum $\hat{S}_{YU}, \hat{S}_{UU}$ is $\hat{G}(k) = \hat{S}_{YU}(k) / \hat{S}_{UU}(k)$ (19). When the measurement is made under feedback conditions (see Figure S8), the output $y(t)$ depends on both the measured input $u(t)$ and the disturbance source $v(t)$. Due to the presence of the feedback loop, the signal u also depends on the disturbance v. As a result, the FRF measurement at frequency k converges to [17]

$$\tilde{G} = \frac{GS_{RR} - \bar{C}S_{VV}}{S_{RR} + |C|^2 S_{VV}}. \tag{S8}$$

This expression reduces to $\tilde{G} = G$ if $S_{VV} = 0$ (r dominates over v), and $\tilde{G} = -1/C$, if $S_{RR} = 0$ (v dominates over r). For mixed SNR, the estimate becomes a mixture of the feedforward and feedback characteristics.

The bias can be eliminated if the external reference signal r is also available [17]. In that case, the indirect method can be used [85], [86]

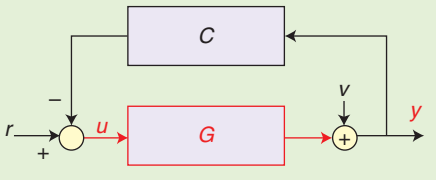


FIGURE S8 The frequency response function (FRF) measurement under closed-loop conditions requires special care. The FRF of the system G captured in a feedback loop is measured starting from the measured input $u(t)$ and output $y(t)$. This leads to a bias because the input u is correlated with the noise v through the feedback path C.

$$\tilde{G} = \frac{G_{yr}}{G_{ur}} = \frac{S_{YR}}{S_{UR}}. \tag{S9}$$

Because r is known exactly and not correlated with v, the bias in (S8) is removed. This approach can also be interpreted within the instrumental variables identification framework [21]–[23].

errors are completely due to leakage errors (see "Models For Dynamic Systems: Finite-Length Effects"), since there was no disturbing noise. Disturbing noise has a very similar effect. For this reason, no distinction is made in further discussions in this section. A detailed analysis of the disturbing noise effects is provided in the section "Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods."

In the absence of input noise, the ETFE is asymptotically unbiased for N, growing to ∞ if the expected value is calculated with respect to the output noise, conditioned on a given realization of the random input [21]. However, its variance does not decrease to zero [21], even if there is no input noise present. To reduce the variance, a proper smoothing procedure is needed. The simple averaging procedure used in the previous section works well for synchronized measurements of periodic data but fails for random noise excitations. For periodic excitations with P measured periods, the sample mean converges to the true value

$$\lim_{P\to\infty} \frac{1}{P} \sum_{l=1}^{P} U^{[l]} = U_0. \tag{15}$$

For random noise excitations that have no periodic nature, it is not possible to split the record over the successive periods. Instead, the original data record is split into P subrecords. In this case, the sample mean $\hat{U}(k), \hat{Y}(k), k \neq 0$

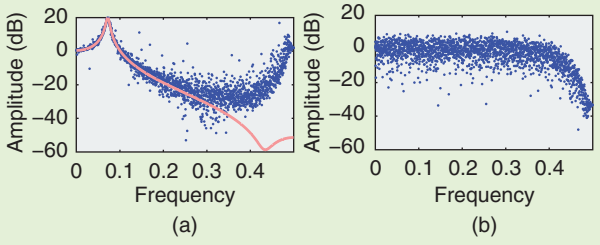


FIGURE 7 The frequency response function measurement shown in Figure S6 (a), together with the amplitude $|U|$ of the discrete Fourier transform of the actual realization of the input that was used for this measurement (b). $|U|$ is a random variable too that varies over the frequency. $|U|$ is very small at some frequencies, resulting in large errors on $\hat{G}$.

converges to zero (that is, $E\{\hat{U}\} = 0$) at the same rate as its standard deviation of the output noise, so that no increase in SNR is obtained. Moreover, the estimate (12) degenerates to 0/0, and hence an alternative averaging procedure is needed. Two approaches are discussed: smoothing over neighboring frequencies and smoothing over successive realizations (subrecords) of the data.

Smoothing the Empirical Transfer Function Estimate over Neighboring Frequencies

If the frequency resolution f_s/N of the FRF measurement is small enough, it can be safely assumed that G does not

Models for Dynamic Systems: Finite-Length Effects

A linear dynamic system is described by (1). When this relation is applied to finite-length records, additional effects appear (as illustrated in Figure S9) that plot the simulated input and output for a random noise excitation. The time domain record is split into three subrecords, and for each of these the corresponding output is calculated (blue, green, and red), setting the input equal to zero outside the subrecord. When only the middle subrecord is processed, it is clear that two problems appear: 1) the beginning of the output measurement is disturbed by the blue output transient of the first subrecord and 2) the green output-transient of the middle subrecord is missing because it belongs to the third output subrecord.

The effect of both transient terms is that the simple relation $Y(k) = G(k)U(k)$ between the discrete Fourier transform (DFT) coefficients $U(k), Y(k)$ of the middle record does no longer hold. An additional term is needed to account for it. Failing to do so creates errors in the impulse response (IR) and frequency response function (FRF) estimates, even for noise-free measurements, as shown in Figure S6.

Adding the transient terms to $Y(k) = G(k)U(k)$ is done below for single-input, single-output (SISO) and multiple-input, multiple-output (MIMO) systems in the time and frequency domains. There is a full equivalence between the time- and frequency-domain equations.

In the time domain, the finite-length effects are known as transient effects due to the additional blue transient in Figure S9. In the frequency domain, the finite-length effects are called leakage errors and are due to the combined effect of the blue and green transient. Usually, leakage errors are considered to be random errors, although it is clear from the previous discussion and the mathematical description below that they are highly structured errors. For a long time, this was not fully realized, and only recently has this insight been fully explored to develop more powerful nonparametric methods that can remove the finite-length effects [19], [20], [66], [67], [69], [71]–[73]. These are discussed in the section "Improved Frequency Response Function Measurements Using Local Parametric Methods."

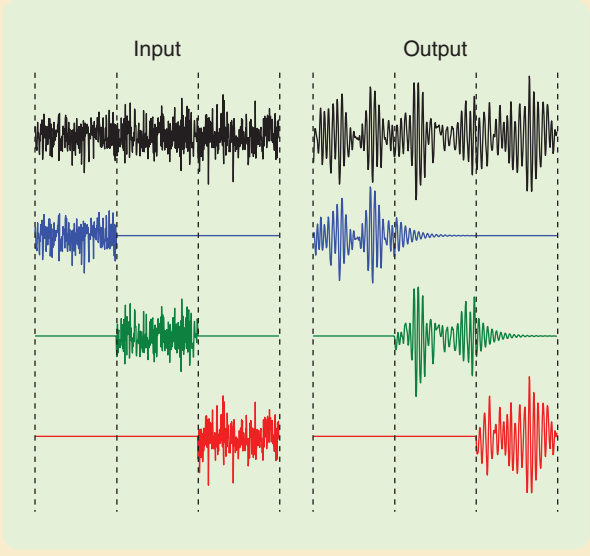


FIGURE S9 Understanding finite-length effects on the relation $Y(k) = G(k)U(k)$ between the discrete Fourier transform coefficients $U(k), Y(k)$ of the middle record for a linear dynamic system. The impact of the beginning (additional blue transient at the start of the middle subrecord) and end effects (missing green transient at the start of the left subrecord) are shown.

SISO SYSTEMS

Time Domain: Including the Transient Effects

Consider the finite-length input/output measurement $u_{FL}(t) = u(t), y_{FL}(t) = y(t), t = 1, \ldots, N$ of the discrete-time SISO system in (1). When (1) is applied to $u_{FL}(t), y_{FL}(t)$, making $u_{FL}(t < 0) = 0$, it needs to be extended with an additional transient term $t_g(t)$ to account for the unmeasured past inputs prior to the start of the experiment

$$y_{FL}(t) = g(t) * u_{FL}(t) + t_g(t). \tag{S10}$$

In a parametric presentation of (S10), the impulse response $g(t)$ and the transient $t_g(t)$ are both modeled by a rational form in the Laplace or z-domain

$$G = B_G/A_G, T_G = I_1/A_G, \tag{S11}$$

with A_G, B_G, and I_1 polynomials in s or z.

Frequency Domain: Modeling the Leakage Errors

Transforming (S10) to the discrete frequency domain using the discrete Fourier transform (DFT) (S1) gives the following relation between the DFT coefficients $U(k), Y(k)$ [16], [19], [70], [71]

$$Y(k) = G(k)U(k) + T_G(k), k = 0, 1, \ldots, N/2. \tag{S12}$$

It can be shown that $DFT\{t_g\} \neq T_G$, because the latter depends on the missing green output transient in subrecord 3 in Figure S9, while this is not the case for $t_g(t)$ [16], [19], [70], [71].

For a random-noise or quasi-stationary excitation [16], [21]

$$U(k) = O(N^0), T_G(k) = O(N^{-1/2}), \tag{S13}$$

with $O(x)$ an arbitrary function with the property (in probability) $\lim_{x \to 0} |O(x)/x| < \infty$. This shows that the transient effect decreases as $O(N^{-1/2})$ to zero for a growing record length.

The parametric presentation (S14) of (S12) models the impulse response $g(t)$ and the transient $t_g(t)$ again by a rational form in the z-domain that is very similar to (S11)

$$G = B_G/A_G, T_G = I/A_G. \tag{S14}$$

However, this time the numerator of the transient term $I = I_1 - I_2$, where the polynomials I_1 and I_2 depend, respectively, on the beginning and end conditions of the finite record. A_G, B_G, I_1, and I_2 are polynomials in z as in the time-domain description. Again, these results can be extended to the continuous-time domain by adding a (small) error term that accounts for the alias errors of the sampled transient term [16].

State-Space Interpretation of the Leakage Errors

An interpretation of these results using an A, B, C, and D state-space representation is also very revealing [71], where

$$T_G(z) = zC(zI - A)^{-1}(I - A^N)(x_0 - x_p), \tag{S15}$$

with

$$x_p = (I - A^N)^{-1} \sum_{t=0}^{N-1} A^t B u(N - 1 - t), \tag{S16}$$

with x_0 the initial state. These expressions show clearly that the additional term is due to the transient response on $x_0 - x_p$, which is the difference in initial conditions between the true system and the assumed periodic one in the DFT analysis. As mentioned before, these expressions are only valid at the DFT frequencies kf_s/N.

The Z-transforms of t_g and T_G are both given by a rational form of the same order, having the same poles as the plant transfer function $G(z)$. These results extend also to continuous-time systems, provided that the sampling frequency is high enough that the aliasing error of the sampled transient signals is negligible.

Experimental Illustration

The above results show that transient and leakage errors are highly structured. The induced errors have a "smooth" aspect, in both the time and the frequency domain. This was a key observation for the development of the new generation of nonparametric frequency-domain algorithms. An experimental illustration of these results is shown in Figure S10.

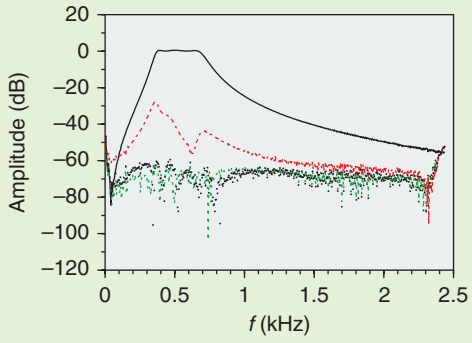


FIGURE S10 An illustration of the importance of the transient term in (S12). A sixth-order bandpass filter is excited with random noise. A reference frequency response function is measured with a periodic, steady-state experiment. Next, the data are used to identify a parametric transfer function model with or without the transient term T_G. The black line is the reference measurement G_{ref} using the steady-state periodic measurements. The red dots are the error on the model without T_G estimated on the random noise data. The green dots are the error on the model with T_G estimated on the random noise data. The black dots are the error of the model that is estimated on the reference data without a transient term. Note that the transient errors are smooth. Including the transient term brings the errors of the full model identified on the random data down to the same level as the errors in the model that are identified from the periodic reference data without the transient term.

MIMO SYSTEMS

All the SISO results can be generalized to MIMO systems with n_u inputs and n_y outputs. These are characterized by the IR matrix $g(t) \in \mathbb{R}^{n_y \times n_u}$ in the time domain, and by the frequency response matrix $G(k) \in \mathbb{C}^{n_y \times n_u}$ in the frequency domain. The transient term becomes a vector $t_g \in \mathbb{R}^{n_y \times 1}$, or $T_G \in \mathbb{C}^{n_y \times 1}$. There are three different choices for the parametric representation of G. The common denominator form

$$G(z) = B(z)/A(z), \tag{S17}$$

with $G \in \mathbb{C}^{n_y \times n_u}$ and A a scalar. Alternatively, the left matrix fraction form

$$G(z) = A(z)^{-1} B(z), \tag{S18}$$

with $A \in \mathbb{C}^{n_y \times n_y}$ and $B \in \mathbb{C}^{n_y \times n_u}$, or the right matrix fraction form

$$G(z) = B(z) A(z)^{-1}, \tag{S19}$$

with $A \in \mathbb{C}^{n_u \times n_u}$ and $B \in \mathbb{C}^{n_y \times n_u}$ can be used. A detailed discussion on the differences and similarities between these choices is out of the scope of this article; more information can be found in [10].

CONCATENATING EXPERIMENTS

A model for concatenated data records is written using the generalized plant plus transient model structure. At each concatenation point, an additional transient is added to the model [74]. This allows longer records to be obtained starting from short experiments and resulting in an increased frequency resolution.

vary much in the frequency interval $B = [k-n, k+n]$, and hence a smoothed estimate for $\hat{\hat{G}}(k)$ can be obtained by calculating a properly weighted average [21], [55], [56]

$$\hat{\hat{G}}(k) = \frac{\sum_B \hat{G}(m)/\sigma_G^2(m)}{\sum_B 1/\sigma_G^2(m)} = \frac{\sum_B \hat{G}(m)|U(m)|^2}{\sum_B |U(m)|^2}, \\ = \frac{\sum_B Y(m)\bar{U}(m)}{\sum_B |U(m)|^2}. \tag{16}$$

The second equality holds true if the output noise dominates (lower SNR at the output than at the input) such that

$$\sigma_{\hat{G}}^2(m) = |G_0|^2 \frac{\sigma_Y^2(m)}{|Y_0^2(m)|} = \frac{\sigma_Y^2(m)}{|U_0^2(m)|}, \tag{17}$$

and the variance $\sigma_Y^2(m)$ remains constant in the interval B.

The major disadvantage of the ETFE smoothing technique is the creation of bias errors (mainly at the resonances and antiresonances) because the estimate $\hat{G}(k)$ is averaged over the frequency interval B. This introduces errors that are proportional to the second derivative of G of order $O(d^2G(f)/df^2)$. For that reason, the width of the smoothing window B should be well tuned. Choosing it too small results in a large variance, while selecting it too large results in a large bias. Optimal tuning methods have been proposed in [57], where an initial estimate of the noise variance $\sigma_Y^2(m)$ is first completed and used to tune the bandwidth B in (16). The proposed method uses local polynomial approximations to estimate the noise variance. This idea will be extended in the section "Improved Frequency Response Function Measurements Using Local Parametric Methods."

In (16) the measurements are again averaged before the division is made. However, $|U|^2$ is averaged in the denominator, so a nonlinear operation is still applied before the averaging. This leads to biased estimates if the input measurement is disturbed by noise. A detailed discussion of the bias and variance effects of the disturbing noise is provided in the section "Bias and Variance Analysis of Frequency Response Function Leakage Errors."

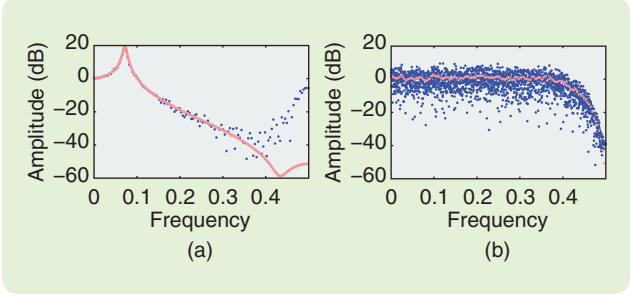


FIGURE 8 The frequency response function (FRF) starting from the simulation in Figure S6. The full record is broken into 16 subrecords of length $N = 256$. (a) Pink line: true FRF G, and blue dots: FRF measurement $\hat{G}(k) = \hat{S}_{YU}(k)/\hat{S}_{UU}(k)$. (b) Blue dots: $|U|$ for the full record, and pink dots: $\hat{S}_{UU}$. The frequency resolution of the pink dots is 16 times smaller than that of the blue dots because the length of the subrecords is 16 smaller than that of the full record.

Smoothing the Empirical Transfer Function Estimate over Successive Realizations

An alternative approach to average the ETFE is to split the original data record into $P = 2n+1$ subrecords, as was done for the periodic excitation solution above. A new estimate is defined by averaging over the subrecords instead of the neighboring frequencies

$$\hat{G}(k) = \frac{\sum_{l=1}^{P} Y(k)^{[l]}\bar{U}(k)^{[l]}}{\sum_{l=1}^{P} |U(k)^{[l]}|^2}. \tag{18}$$

Smoothing over the frequency interval B is avoided, but the frequency resolution of the subrecords is P times smaller than that of the original record and equals the width of the interval B. Hence, similar bias errors will appear for (18) as for (16) (see also the section "Bias and Variance Analysis of Frequency Response Function Leakage Errors" and Table 2).

The scaled numerator and denominator in (18) can also be interpreted as estimates of the cross spectrum S_{YU} and auto spectrum S_{UU} between u, y, so that (18) can be rewritten as

$$\hat{G}(k) = \frac{\hat{S}_{YU}(k)}{\hat{S}_{UU}(k)}, \tag{19}$$

which links the ETFE method directly to the classical FRF estimation methods of the 1960s [1], [2], [11], [17], [58], [59] that will be discussed in detail in the next section. The stochastic properties are also analyzed in the next section.

In Figure 8, the data from Figure S6 are processed using (19). The original data record of 4096 samples is split into 16 subrecords of 256 samples each, and (18) is then calculated. Averaging over 16 subrecords significantly reduced the presence of the dips in S_{UU} [see Figure 8(b)] with respect to $|U|$, resulting in a smoother FRF estimate as shown in Figure 8(a) when compared to previous results using a single record (as shown in Figure 7). This was at a cost of a reduced frequency resolution. The FRF is measured at 128 frequencies instead of 2048 frequencies in (8).

User Guideline

The spectral resolution is $\Delta_f = BP/N$, with B the smoothing width, P the number of subrecords or repeated experiments, and N the total data length. The noise reduction is of $O(1/\sqrt{BP})$. The number of subrecords P and the smoothing bandwidth B can be interchanged without affecting the effective frequency resolution or the uncertainty of

the result if the disturbing noise is dominating. The leakage errors will dominate for a high SNR, in which case it is better to increase B and keep P as small as possible. This is because, for a given length N, the impact of the leakage errors grows with $\sqrt{P}$.

Smoothing the Frequency Response Function Using the Classical Spectral Estimation Methods

For a long time, an FRF measurement was the division of two spectra [1], [2], [4], [11], [18] [see (12), (19)]. Hence, the individual spectra U,Y had to first be accurately estimated, instead of determining a model between these two signals. This was a more demanding approach that was more sensitive to leakage errors. Measuring the spectrum $X(k), k = 0,\ldots,N/2$ of a random signal is a tedious job. Due to its random nature, it is not straightforward to obtain a high-quality measurement of the spectrum starting from a finite-length measurement $x(t), t = 1,\ldots,N$. The DFT (S1) $X(k)$ of $x(t)$ is not a consistent estimate because its variance does not decrease to zero for a growing length N. This is because the spectral resolution is growing with N, such that no "averaging" occurs, as is shown in Figure 9. For this reason, it is necessary to explicitly introduce an averaging step in the estimation procedure to obtain a reliable spectral estimate. This can be accomplished over neighboring frequencies or multiple realizations (see [11] and the references therein), similar to the previous discussion of the ETFE. Eventually, the latter approach became the dominant one, especially after the introduction of the FFT [34] to calculate the DFT [60]. The main motivation was faster computation and reduction of the core storage necessary to process the shorter records [2].

These different approaches to experimental spectral analysis were also reflected in different proposals for FRF estimators using noise excitations. The standard procedure (known as Welch's method [2]) or the weighted-overlapped segment averaging procedure splits the original long data record $u(t),y(t)$ into P shorter (overlapping) subrecords $u^{[l]},y^{[l]}, l = 1,\ldots,P$ of length N and calculates for each of these the DFT $U^{[l]},Y^{[l]}$ (see Figure 10). The auto spectrum S_{UU} and cross spectrum S_{YU} are estimated using the expressions

$$\hat{S}_{UU}(k) = \frac{1}{P}\sum_{l=1}^{P} \left| U(k)^{[l]} \right|^2,$$
$$\hat{S}_{YU}(k) = \frac{1}{P}\sum_{l=1}^{P} Y(k)^{[l]}\bar{U}(k)^{[l]}. \tag{20}$$

These are equal to the numerator and denominator in (18), which clearly links both methods. In Figure 11, it is illustrated that averaging over P realizations using these expressions reduces the variability. The dips disappear even more rapidly than would be expected with the $1/\sqrt{P}$ rule (see "Impact of Averaging on the Variance of the Frequency Response Function Estimate"), especially for low values of P.

The FRF $\hat{G}(k)$ is then obtained using (19). It is essential in this approach to realize that $\hat{G}(k)$ is obtained as the division of two estimated spectra that are prone to errors due to the finite data length effects (called leakage errors) and disturbing noise. The errors in Figure 8 are

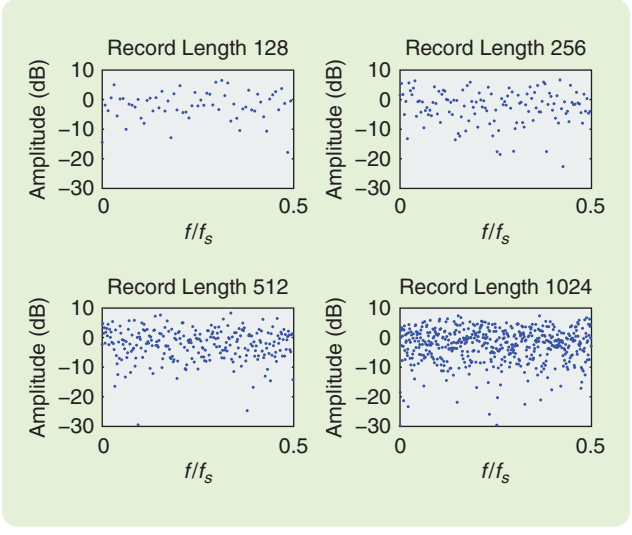


FIGURE 9 The discrete Fourier transform spectrum of white noise for a growing record length. The spectrum does not converge to a constant, as it could be expected for white noise. Longer measuring does increase the number of frequencies on which the spectrum is calculated, but the variability of the measurement does not decrease.

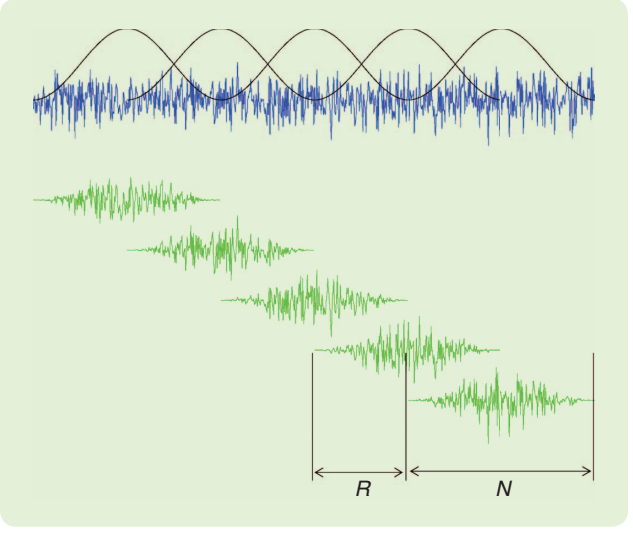


FIGURE 10 An illustration of the overlapped weighted segments—here with a Hanning window w_H (black lines) and an overlap $1 - R/N = 1/2$ (the window is shifted R samples in each step). Each subrecord is multiplied with the window (shown in green, shifted below the original signal in blue), and the discrete Fourier transform is applied to the product.

These different approaches to experimental spectral analysis were also reflected in different proposals for FRF estimators using noise excitations.

completely due to the leakage; there was no disturbing noise added to the data.

The majority of the previous efforts focused on a better understanding and tuning of the leakage errors using well-adopted time windows [4] that reshaped the errors to become less harmful. The impact of both the leakage and noise errors will be analyzed in more detail in this section, because their combined effect determines the stochastic properties of $\hat{G}(k)$. Note that the alias errors are not considered anymore; it is assumed that the antialias filters and the sampling frequency are properly matched, so that these errors can be neglected.

Time- and Frequency-Domain Interpretation of Windows

Calculated DFT spectra are prone to leakage errors due to finite length effects [34]. A subrecord is obtained from an infinitely long signal by multiplying the original signal with a "selection window" $w(t)$ that is equal to zero outside the selected time interval. For example, as shown in Figure 10, the lth subrecord is $u^{[l]}(t) = w^{[l]}(t)u(t)$. Multiplication in the time domain becomes a convolution in the frequency domain, so that

$$U^{[l]}(k) = W(k) * U^{[l]}(k). \tag{21}$$

Many windows have been discussed in the literature, each being optimized for a given application (such as spectral resolution, detecting small spurious frequency components, and amplitude measurements [61]). In this article, the rectangular, Hanning, and Diff windows will be considered. However, instead of studying their impact on the spectral estimates U, Y [3], the focus will be directly on the impact of windowing on the relation (S12)

$$Y(k) = G(k)U + T_G(k), \tag{22}$$

following the approach in [19], [62], and [63]. All the windows $w(t)$ are defined on the half-open interval $t \in [0,1)$ for $t = [0,1,\ldots,N-1]/N$ [see Table 1; outside this interval $w(t) = 0$)]. The definition is given in the time domain, and the impact of the window on the DFT spectrum is given in the last column. The Diff and the half-sine windows are shown to have very similar properties for FRF measurements [63].

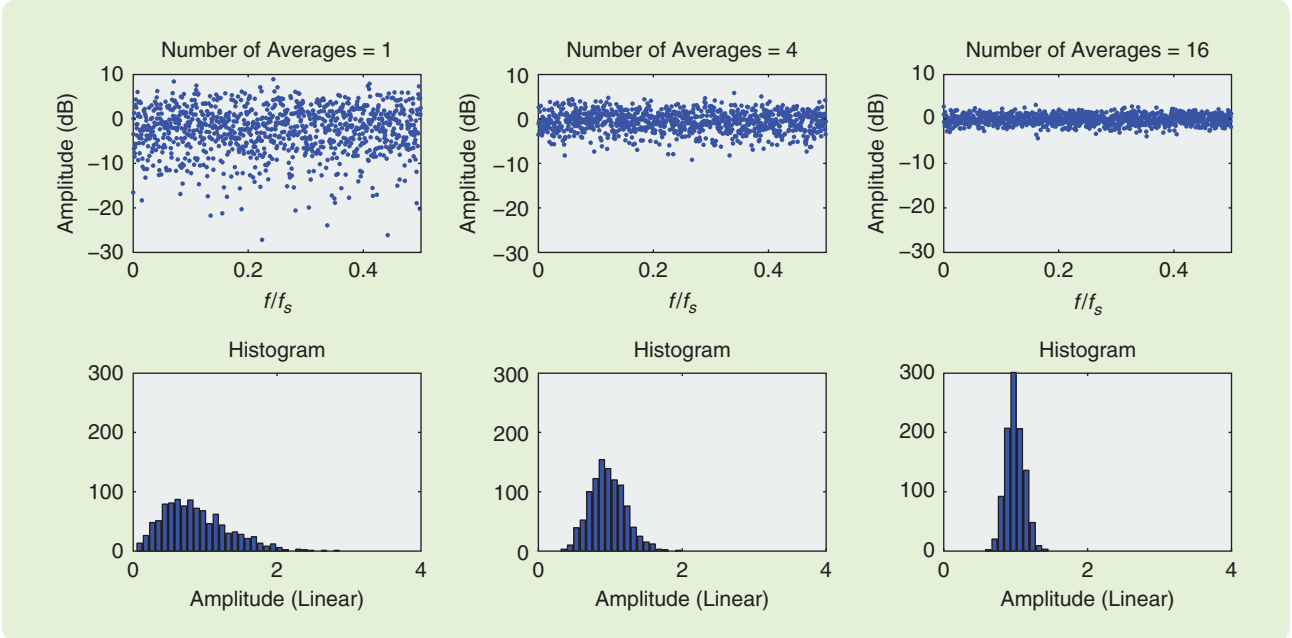


FIGURE 11 An illustration of the smoothing effect obtained by averaging $\hat{S}_{UU}$ over different realizations. The power spectrum is shown in the upper figures for $M = 1, 4, 16$ and $N = 2048$. The lower figures show the corresponding histograms. It can be shown that these follow a χ^2 distribution with $2P$ degrees of freedom [16].

The Diff window $X_w(k+1/2) = [X(k+1) - X(k)]$ acts as a difference centered around $k+1/2$, where the latter frequency index points to the frequency $f = (k+0.5)f_s/N$. The Hanning window $X_w(k) = \frac{1}{4}[-X(k-1) + 2X(k) - X(k+1)]$ is a double difference centered around k. From this observation, it can be easily understood that windows push down spectra that are smooth in function of the frequency, as is the case for the transient $T_G(k)$ in (S12). The first term in this expression $G(k)U(k)$ will be "rough" for random noise excitations and hence not be reduced by the windowing operation. This is also illustrated in Figure 12, where the amplitude spectrum of the rough signal and the smooth transient term are plotted using a rectangular and a Hanning window. This shows that windowing can decrease the leakage errors, since these are directly coupled to the transient term T_G.

Understanding the Impact of Leakage on Frequency Response Function Measurements

Windowing will not remove all the errors; the tapered output will not be identical to the response of the system on the tapered input, as was previously shown in Figure S9. The precise impact on the FRF measurement is studied in this section. To keep the focus completely on the leakage errors, it is assumed that the disturbing noise is zero, so the errors in the FRF measurement will be solely due to leakage effects.

Assumption 4

No disturbing noise

$$U(k) = U_0(k),$$
$$Y(k) = Y_0(k). \tag{23}$$

The impact of the Hanning window on the FRF measurement is obtained by replacing $Y = GU + T_G$ at the three frequencies $k-1$, k, and $k+1$ in $Y_w(k) = \frac{1}{4}[-Y(k-1) + 2Y(k) - Y(k+1)]$. The smooth functions $G(O(N^0))$, and $T_G(O(N^{-1/2}))$ can be approximated by [62]

$$G(k \pm 1) = G(k) \pm \triangle_G + O(N^{-2}),$$
$$T_G(k \pm 1) = T_G(k) \pm \triangle_{T_G} + O(N^{-5/2}), \tag{24}$$

with $\triangle_G = O(N^{-1})$. This results in

$$Y_w(k) = G(k)U_w(k) + E_G(k) + E_{T_G}(k), \tag{25}$$

with

$$E_G(k) = -\triangle_G(U(k+1) + U(k-1)) = O(N^{-1}),$$
$$E_{T_G} = O(N^{-5/2}). \tag{26}$$

The transient error E_{T_G} remains from the transient after the double differentiation. $E_G(k) = O(N^{-1})$ is a new error that is due to interpolation of G over the left and right frequency in the Hanning window. Since it became the dominating error, it can be concluded that the Hanning window replaces the leakage error that is of $O(N^{-1/2})$ by a smaller $O(N^{-1})$ interpolation error. These results are tabulated in Table 2 [63] and compared to those of the rectangular and Diff (half-sine) windows. The results in Table 2 are normalized by the time constants of the system [64]. To do so, the dominant time constant τ_f in the frequency b and of interest is selected, and N is replaced by N/τ_f. For a resonating system, the dominant time constant is set by the damping of the actual resonance that is studied (see "Characterizing a Resonance by Its 3-dB Bandwidth").

Since the rectangular window makes no interpolation, it follows that only the leakage error will be present

$$Y_R(k) = G(k)U_R(k) + 0 + E_{T_G}. \tag{27}$$

For the Diff window with a width of one bin, the same expression as the Hanning window holds. The latter has a width of two bins, so the interpolation error for the Diff window will be smaller [although it is also $O((N/\tau_f)^{-1})$,

TABLE 1 Many windows are defined and studied in signal processing [61]. Popular windows include the rectangular and Hanning windows. Recently, the Diff and half-sine windows were studied in detail [63]. The definition of these windows $w(t)$ is given for $t \in [0, 1)$.

Window	$w(t)$	X_w
Rectangular	1	$X_w(k) = X(k)$
Hanning	$0.5(1 - \cos 2\pi t)$	$X_w(k) = \frac{1}{4}[-X(k-1) + 2X(k) - X(k+1)]$
Diff	$e^{j2\pi t} - 1$	$X_w(k+1/2) = [X(k+1) - X(k)]$
Half-sine	$\sin \pi t$	$F(\sin \pi x)$

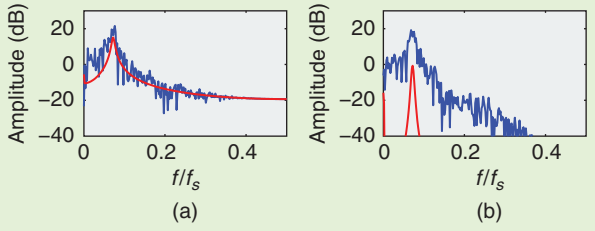


FIGURE 12 The impact of the (a) rectangular and (b) Hanning window on the signal term GU (blue) and the transient term T_G (red). The signal term is "rough," and the transient term is smooth. The double differentiation of the Hanning window reduces the smooth transient term, while the amplitude level of the rough signal term remains the same. The system is the same as in Figures S6 and 8, and the data length is $N = 512$.

as it was for the Hanning window]. This is still larger than the leakage error of the Diff window, which is $E_{T_G} = O((N/\tau_f)^{-3/2})$, because only one difference is made.

Remark 2

Exponentially decaying windows are also used [65] in modal analysis (analysis of vibrating mechanical structures) [18]. These add an artificial but known damping to data that can be compensated for when physical parameters are extracted from the FRF data.

Bias and Variance Analysis of Frequency Response Function Leakage Errors

For random excitations, the leakage error is itself a random variable. In "Models for Dynamic Systems: Finite-Length Effects," it is explained that leakage errors are created by transient effects at the output of the system and these have a smooth spectrum $T_G(k)$. The leakage error is then given by $T_G(k)/U(k)$, and, since $U(k)$ is a random variable, the leakage error will be too. This error is conditioned on the random input, and it is characterized by its mean value (bias errors) and variance [4].

Bias Errors

Averaging over successive realizations of the input reduces this error, but it will not decrease toward zero as the number of averages P approaches ∞. We note that the leakage error is due to the beginning and end transient. These depend linearly on the random input signal, and hence the correlation $E\{T_{T_G}U\}$ between the input and the transient will be different from zero, leading to systematic errors (bias) in the FRF measurement [66]. For the rectangular window, the bias error is shown to be $O((N/\tau_f)^{-1})$ [62], [63]. For the Hanning and the Diff (half-sine) windows, the bias error is mainly due to the interpolation errors and drops to $|G^{(2)}(k)O((N/\tau_f)^{-2})$. The bias error of the Diff window is slightly smaller than that of the Hanning window, because the width of the Diff window is half that of the Hanning window.

TABLE 2 Applying a window in frequency response function measurements reduces the transient errors E_{T_G} and creates interpolation errors E_G. These errors, together with their bias and variance, are given as a function of N: the length of the subrecord, the dominant time constant in the frequency band of interest τ_f, and the second derivative of G with respect to the frequency $G^{(2)}$.

Window	E_{T_G}	E_G	Bias	Variance
Rectangular	$O((N/\tau_f)^{-1/2})$	0	$O((N/\tau_f)^{-1})$	$O(P^{-1}(N/\tau_f)^{-1})$
Hanning	$O((N/\tau_f)^{-5/2})$	$O((N/\tau_f)^{-1})$	$\lvert G^{(2)}(k)\rvert O((N/\tau_f)^{-2})$	$O(P^{-1}(N/\tau_f)^{-2})$
Diff or half-sine	$O((N/\tau_f)^{-3/2})$	$O((N/\tau_f)^{-1})$	$\lvert G^{(2)}(k)\rvert O((N/\tau_f)^{-2})$	$O(P^{-1}(N/\tau_f)^{-2})$

Variance Errors

Prior literature [2], [4] focused on the bias contribution of leakage. Little effort was spent on the

Characterizing a Resonance by Its 3-dB Bandwidth

The transfer function of a system can be written as the sum of first-order subsystems (with a real pole) and second-order subsystems (with two complex conjugate poles). Lowly damped poles create (sharp) resonances in the frequency response function (FRF), and many FRF methods are most prone to errors around these resonance frequencies, as shown in Figure 14.

The frequency resolution δ_f of the FRF measurement should be "small" enough to properly capture the resonance peaks; a too-small frequency resolution would underestimate the peak value of the FRF. The 3-dB bandwidth of a resonance can be used to make a quantified statement. In Figure S11, the 3-dB bandwidth is shown in pink. The 3-dB bandwidth $B_{3\text{dB}}$ in rad/s is given by the width of the frequency interval around the peak amplitude of the FRF, where $|G(\omega)|_{\text{dB}} \geq G_{\max} - 3$.

There is a direct link between $B_{3\text{dB}}$, the damping ζ, and the time constant τ of a second-order system with resonance frequency ω_n

$$B_{3\text{dB}} = 2\zeta\omega_n = 2/\tau. \tag{S20}$$

The frequency resolution of an FRF measurement should be high enough to cover all resonances of interest. For example, the number of frequency points $F_{3\text{dB}}$ in the 3-dB bandwidth where the FRF is measured is $F_{3\text{dB}} = N_{\text{sub}}/(\pi\tau f_s)$, which directly relates the time constant τ of the system to the record length N_{sub}.

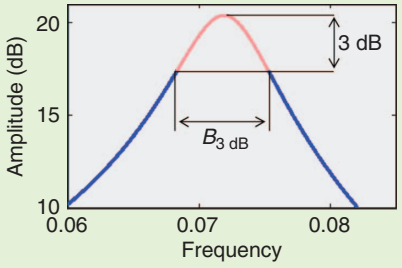


FIGURE S11 Characterizing a resonating system by its 3-dB bandwidth, indicated in pink.

study of leakage-induced variance, although under good SNR conditions, the latter becomes the dominating error. The variance error is respectively of $O(P^{-1}(N/\tau_f)^{-1})$ for the rectangular window and $O(P^{-1}(N/\tau_f)^{-2})$ for the Diff (half-sine) and Hanning windows.

Overlapping Windows

While it is not possible to reduce the bias, the variance can be further reduced by allowing for an overlap of R samples, when shifting the window with length N over the long record [63], [67]. This increases the calculation time (because more subrecords need to be processed), but the variance is further reduced. Because in most applications, the measurement time is a more important concern than the calculation effort, the overlapping strategy is the standard procedure in current practice. Because the overlapping subrecords are more correlated for an increasing overlap, the gain saturates. In [63], a very detailed study shows that, under very loose conditions, the optimal choice among all windows with a bounded derivative is the Diff (half-sine) window with an overlap of $1 - R/N = 2/3$ (the window is shifted each time with $N/3$). This results in a further reduction of the leakage-induced variance by more than a factor of 3.5 (half of this at the zero and half at the sample frequency). This result will also extend to the disturbing noise sensitivity. However, the gain is about a factor of two.

Variance Analysis of the Frequency Response Function Measurements in the Presence of Disturbing Noise

The variance of $\hat{G}$ (19) due to the disturbing input and output noise N_U, N_Y can again be retrieved by linearizing the expression with respect to the noise, similar to the approach in (12), but is now applied to the estimates $\hat{S}_{YU}(k)$ and $\hat{S}_{UU}(k)$. This results eventually in

$$\sigma^2_{\hat{G}} = \frac{1}{P}|G_0|^2\left(\frac{\sigma^2_Y}{\hat{S}_{YY}(k)} + \frac{\sigma^2_U}{\hat{S}_{UU}(k)} - 2\,\mathrm{Re}\left(\frac{\sigma^2_{YU}}{\hat{S}_{YU}(k)}\right)\right). \tag{28}$$

For P approaching ∞, the estimated auto and cross power spectra can be replaced by their exact value, and it is clear that $\sigma_{\hat{G}}$ drops again in $1/\sqrt{P}$. However, for smaller values of P, this approximation is no longer valid (due to the presence of dips as illustrated in Figure 11), and the original expression (28) should then be used. The additional loss is plotted in Figure 13 [16], showing that, for small values of P, the excess loss is close to five for $P = 1$. Using overlapping subrecords, the value of P is artificially increased by a factor of two, as explained in the previous section.

A standard alternative for the variance expression (28) is [4], [18]

$$\sigma^2_{\hat{G}(k)} = \frac{1}{P}|G_0(k)|^2\frac{1-\gamma^2(k)}{\gamma^2(k)}, \tag{29}$$

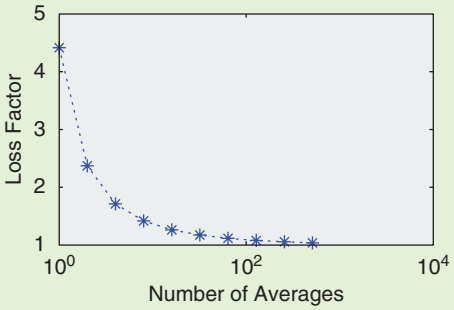


FIGURE 13 Th impact of replacing S_{UU} by its finite sample estimate $\hat{S}_{UU}$ in $\hat{G} = \hat{S}_{YU}/\hat{S}_{UU}$. The estimated power spectrum $\hat{S}_{UU}$ is a random variable that can become very small when the number of averages is small, as shown in Figure 11. These dips create large spikes after the division in $\hat{G}$, leading to an increase of the standard deviation $\sigma_{\hat{G}}$ of $\hat{G}$. The 95% bound of the relative increase of the standard deviation $\sigma_{\hat{G}}$ is plotted.

with

$$\gamma^2(k) = \frac{|S_{YU}(k)|^2}{S_{UU}(k)S_{YY}(k)}. \tag{30}$$

In practice, the coherence γ^2 is estimated by replacing the theoretical values in (30) by their measured values. The coherence (30) has exactly the same interpretation as the correlation (11) for periodic excitations [4]. The coherence is $\gamma^2 = 1$ for undisturbed measurements (no leakage, no noise, no nonlinear distortions), pointing to perfectly linearly related data. It drops to zero due to noise disturbances, unmeasured inputs, and data that are not linearly related [4]. It is a very popular measure for the quality of the data, and it is plotted in many commercial signal analyzers [3], [4]. Compared to the variance analysis for periodic data (10), the coherence provides less information because it makes no split between input and output noise and their mutual correlation. This can be seen in the analysis of the flexible robot arm data in Figure 6, where the coherence (correlation) is plotted in (c). If only this information is available, the insight in the origin of the dips in the coherence would be lacking (a drop in SNR of the input).

Bias Analysis of the Frequency Response Function: Impact of Noise Disturbances on the Reference Input and the H_1, H_2 Methods

The most popular FRF estimate $\hat{G}$ is (19). However, this method fails completely in the presence of input noise. Consider

$$\lim_{P\to\infty}\hat{G}(k) = G(k)\frac{1+\dfrac{\sigma^2_{YU}}{\hat{S}_{YU}(k)}}{1+\dfrac{\sigma^2_{UU}}{\hat{S}_{UU}(k)}}. \tag{31}$$

Assume for simplicity that the noise is not correlated with the input u and the output y. It follows that (31) equals

$$\lim_{P\to\infty}\hat{G}(k) = G(k)\frac{1}{1+\dfrac{\sigma_{UU}^2}{\hat{S}_{UU}(k)}}, \tag{32}$$

proving that (19) underestimates the true value of the FRF in the presence of noise on the input signal [4], [16]. In the mechanical community, this estimator is called the H_1 estimator.

There is a simple trick to address this problem, provided that the SNR of the output is (very) high. In that case, the H_2 method can be used [68]

$$\hat{G}(k) = \frac{\hat{S}_{YY}(k)}{\hat{S}_{YU}(k)}. \tag{33}$$

The squared input $|U|^2$ is no longer averaged, and the source of the bias in (19) is eliminated. As a rule of thumb, the reference should always be selected as the signal with the highest SNR, resulting in the H_1 or H_2 method [4], [18]. Commercial dynamic signal analyzers use the H_1 method. Switching the input and output cables transforms it into the H_2 method by inverting the estimated FRF $1/\hat{G}$.

User Guidelines

- *Periodic or random excitation*: The first advice remains to use periodic excitations whenever it is possible. If this is not possible for technical reasons, the most recent and powerful FRF measurement techniques in the next section should be used. If, for some reason, the classical methods in this section should be used, the following advice can help to obtain the best results within the classical framework.
- *Bias errors*: Check the experimental setup, and verify the SNR of the input signal. If it is well above 40 dB (1% noise floor), the relative bias of the H_1 will be below 10^{-4}. For a lower SNR, the bias should not be too large for the application. If the bias is too large, it might be an option to switch to the H_2 method if the output has a higher SNR than the input.
- *Variance errors*: It is necessary to average over a sufficiently large number ($P > 16$) of subrecords to keep the additional variance loss small. Use overlapping subrecords (66% overlap) to further reduce the variance.
- *Leakage errors 1*: Even for large SNR levels, the FRF estimate can still be poor due to leakage errors. Use the half-sine (Diff) window in combination with averaging over subrecords to reduce the error. Keep the subrecords as long as possible, because the leakage errors (bias, variance) drop in $1/N^2$. We note that leakage creates bias and variance errors.
- *Leakage errors 2*: The authors strongly advise use of the methods in the next section, because these almost completely eliminate the leakage errors.

IMPROVED FREQUENCY RESPONSE FUNCTION MEASUREMENTS USING LOCAL PARAMETRIC METHODS

Leakage errors were long considered to be random errors. Only recently did it become clear that there is plenty of structure in these errors [66], [67], [69]–[72], as discussed in "Models for Dynamic Systems: Finite-Length Effects," leading to a new family of FRF methods [19], [20], [70], [73], [74]. These recently developed methods do not target a precise estimation of the cross and auto spectrum as the classical methods did, but focus directly on the estimation of the relation (S12), $Y(k) = G(k)U(k) + T_G(k)$ between the DFT spectra. This almost completely eliminates the leakage error, so that only the noise errors remain important. Estimation of $G(k)$ considers a narrow frequency interval $B = [k-n, k+n]$ around the frequency of interest k. In that interval, the parametric models $G = B_G/A_G, T_G = I/A_G$, (S14) can be represented by low-order approximations, so that at each frequency k, a reduced system identification problem is solved. These methods will be shown to have better leakage reduction at a higher computational cost, while the noise sensitivity remains very similar to the classical windowing methods of the previous section.

A short discussion and comparison of the properties of the three local parametric models will be given. The mathematical details, proofs, and more extensive comparisons are discussed in [20], [64], [75], and [76].

The use of a local parametric model conflicts with the concept of nonparametric methods. However, compared to a full parametric approach, identifying a local parametric model is much simpler and requires no user interaction. This is because the model order will be fixed, and the (nonlinear) optimization problem is easily solved due to the simple nature of the local approximation problem. Moreover, the number of local models to be identified grows with the number of data. For these reasons, the local parametric methods still belong to the class of nonparametric methods.

The Local System Identification Problem

A system identification problem is defined by three players: the data, the model, and the cost function. In this case the data are given by (7)

$$\begin{aligned} U(k) &= U_0(k) + N_U(k), \\ Y(k) &= Y_0(k) + N_Y(k), \end{aligned} \tag{34}$$

in the interval $B = [k-n, k+n]$. The local parametric model that is valid in the interval B is

$$Y_0(k) = G(k, \theta_G)U_0(k) + T_G(k, \theta_{T_G}). \tag{35}$$

Both the transfer function and transient term in this model are estimated by minimizing the errors $E(k)$ on the frequency

interval $B = [k-n, k+n]$. The width $2n$ of this interval is called the local bandwidth of the method

$$E(k) = Y(k) - G(k)U(k) - T_G(k), \tag{36}$$

using a weighted least-squares cost function

$$V(k) = \sum_B W(l)\left|E(l)\right|^2 \tag{37}$$

that is minimized with respect to θ_G and θ_{T_G}. For a sufficiently small frequency interval B, the (co)variances of the noise N_U, N_Y can be assumed to be constant, so that these do not affect the optimal choice for the weighting $W(l)$ in (37). Instead, the combined choice of the weighting and the model structure will be used to manipulate the complexity of the identification problem.

Iterative Local Rational Method

$W(l) = 1,\ G = B_G/A_G, T_G = I/A_G.$

This setting results in

$$V_{ILRM}(k) = \sum_B \left| Y(l) - \frac{B_G(l)}{A_G(l)}U(l) - \frac{I(l)}{A_G(l)}\right|^2. \tag{38}$$

The presence of the denominator $A_G(l)$ leads to a nonlinear optimization problem that should be solved iteratively, which is affordable nowadays.

Properties

From the three proposed methods, the iterative local rational method (ILRM) seems to be the most natural choice. This method suffers from a higher noise sensitivity than the other two methods. If the order is not well tuned, excess poles and zeros can create very large narrow spikes due to very closely spaced poles and zeros. This can only be avoided by a dedicated model tuning at every frequency, which reduces the robustness of this method even more. These effects are illustrated in [75, Sec. 3.4].

Local Polynomial Method

$W(l) = 1,\ G = B_G, T_G = I.$

The nonlinear optimization problem is turned into a linear least-squares problem by setting the denominator equal to one $A_G(l) = 1$, which leads directly to the local polynomial method [19], [77],

$$V_{LPM}(k) = \sum_B \left| Y(l) - B_G(l)U(l) - I(l)\right|^2. \tag{39}$$

The minimizer of the cost function is found by solving a linear set of equations, hence no iterative procedure is needed any more.

The pole/zero cancellation problems of the ILRM method are completely eliminated by putting the denominator equal to one. This results in a local polynomial approximation of the transfer function G and the transient T_G. This simplified approach turns out to be very attractive. Besides reducing the optimization problem to a linear one, it also makes the model selection problem less critical. The order R of both polynomials, B_G and I, can be set equal to $R = 2$ (or $R = 4$) with good results. Under these conditions, the disturbing noise induced variance of the local polynomial method (LPM) is 1.74 dB below that of the classical Hanning method if both methods are tuned to the same frequency resolution.

It is clear that the polynomials can only approximate a rational form in the finite frequency window B. The approximation errors are studied in full detail in [64]. The main conclusion is that it is most efficient to choose R to be even. For that choice, the leakage error E_{LPM} is bounded by

$$E_{LPM} = O((B/B_{3\text{dB}})^{R+1}) = O((NP)^{-(R+1)}), \tag{40}$$

with $B_{3\text{dB}}$ the 3-dB bandwidth of the resonance under study. Compared to the windowing methods that had errors of $O(N)^{-1}$, a huge gain is made in the reduction of the leakage errors. From (40), it follows also that the local bandwidth $2n$ of the interval B should be chosen as small as possible, that is, more than two times smaller than $B_{3\text{dB}}$. At the same time, it should contain enough frequencies to estimate the $2(R+1)$ complex coefficients in the two polynomials, so that $n \geq R+1$. This leads for $R = 2$ and $n = 3$ to at least seven frequencies in the interval B.

The last result can also be translated in the minimum record length that is needed [64]. Using the relations in "Characterizing a Resonance by Its 3-dB Bandwidth," $B \leq B_{3\text{dB}}/2$, and that $\tau = 1/(\zeta\omega_n) = 2/B_{3\text{dB}}$, it is found that the frequency resolution of the measurement should be better than $B_{3\text{dB}}/(2R+2)$, and the record length

$$T_{\text{meas}} \geq (R+1)\pi\tau. \tag{41}$$

Local (Linear) Rational Method

$W(l) = \left|A(l)\right|^2, G = B_G/A_G, T_G = I/A_G.$

An alternative to linearize the cost function (38) is to put the weighting $W(l) = \left|A_G(l)\right|^2$, resulting in another linear least-squares problem [78], [79]

$$V_{LRM}(k) = \sum_B \left| A_G(l)Y(l) - B_G(l)U(l) - I(l)\right|^2. \tag{42}$$

Properties

The LPM was the start of a new era in FRF measurements. It was the first method that was proposed and studied in full detail to remove the leakage errors almost completely. However, for systems with low damping (as often happens in advanced mechanical applications), it can be hard to meet the constraint that $B \leq B_{3\text{dB}}/2$ or alternatively that $T_{\text{meas}} \geq (R+1)\pi\tau$. In that case, the LRM approach [79] can still solve the problem because it still identifies a rational

model that can deal with these lowly damped poles. For $R = 2$ (a second-order model), the error of the LRM is $O((B/D)^4)$, with D the shortest distance to the neighboring poles and zeros.

The LRM combines the advantages of the LPM (linear in the parameters) and the IRLM (a rational model). However, for low SNR of the input, a relative bias of $O(\sigma_Y^2 / S_{Y_0 Y_0})$ appears [75]. This is not a real issue in most practical problems, and for that reason the authors advise the LRM method as the default choice among the discussed local parametric models.

General Remarks

History

The LPM was the first local parametric method proposed in the literature [19], [20], and a detailed discussion of its properties for SISO and MIMO can be found in [20] and [77]. In 2012, the LRM method [79] was proposed as an attractive alternative that can better deal with lowly damped systems as often appear in vibrating mechanical structures. The ILRM method is studied in detail in [75]. It is the most expensive method since an iterative algorithm is needed, and it will turn out to be most sensitive to noise among the three proposals.

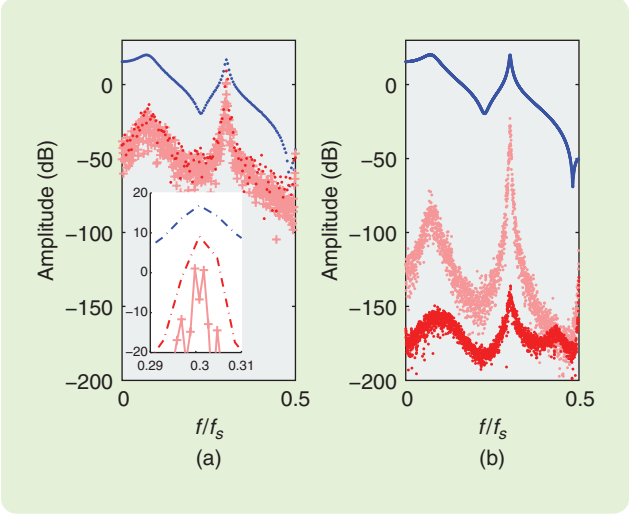


FIGURE 14 An illustration of frequency response function (FRF) measurements using the classical Hanning method and the local polynomial method (LPM) and local rational method (LRM) on undisturbed data. A system with a highly and a lowly damped resonance is excited with filtered white noise (bandwidth $0.4\,f_s$). The full record length is 4096 samples. The blue dots are the estimated FRF, and the red and pink dots plot the errors. Part (a) shows the results for the Hanning method that is applied on subrecords with a length of $N = 256$ (red) and $N = 1024$ (pink) samples, both with an overlap of $R = 2N/3$. In the inset of (a), a zoom around the second resonance is given. Part (b) shows the results for the LPM (pink) and LRM (red), applied to the full length record. The errors of the LPM and LRM are an order of magnitude smaller than those of the Hanning method. The LRM method outperforms the LPM method.

Alternative Methods

Alternative parametric approaches are discussed in [76] and [80]. The first alternative results in a "global" method that links all frequencies to each other, leading to large sets of equations to be solved. The second method is a Bayesian approach that makes an intrinsic tradeoff between variance and bias (see "Bias and Variance Tradeoff of Estimators"). These aspects are discussed later in this article in more detail.

Automatic Tuning of the Local Bandwidth B

Reference [57] proposes an automatic local bandwidth tuning algorithm that starts from a local polynomial model. These ideas can be transferred to the local parametric methods. A first attempt to do so is presented in [81]. A lower root mean square (RMS) error can be obtained at a cost of additional calculations.

Missing Input and Output Data

If data are lost due to sensor failure, overloads, and/or data transmission errors, special actions are needed to address these errors. Instead of making new measurements, advanced signal processing methods can be used. The missing data are then estimated together with the FRF and its variance. If the reference signal is available, missing data in the input and the output can be restored. If that is not the case, the methods assume that only output data are missing [82], [83].

Illustration of the Leakage Rejection of the Hanning, Local Polynomial, and Local Rational Methods

The classical window method (Hanning) and the local parametric methods LPM/LRM are illustrated on a system with two resonances using noise free data (no disturbing noise added), so that the effect of the leakage errors is clearly visible (see Figure 14 for more details). For the Hanning method in Figure 14(a), subrecords are used with a length of $N = 256$ (red) and $N = 1024$ (pink) samples, both with an overlap of $R = 2N/3$. Observe that the errors of the Hanning method become very large, especially around the second resonance (30% or more for the short subrecord length). Using longer subrecords reduces this error, but at a cost of fewer averages (larger risks for spiky errors due to dips in the input spectrum $\hat{S}_{UU}$). In Figure 14(b), the results of the LPM (pink) and LRM (red) are shown. For both methods, the errors are an order of magnitude smaller than those of the Hanning method. Around the second resonance, the error of the LPM increases sharply, because the local bandwidth B becomes large with respect to the 3_{dB} bandwidth. The LRM method performs very well under these conditions.

User Guidelines: Classical Window Methods—Local Polynomial and Local Rational Methods

» *Random excitation*: For measurements with a high SNR of the input measurements, the LPM/LRM provide results with a comparable quality as those

The LRM combines the advantages of the LPM (linear in the parameters) and the IRLM (a rational model).

obtained using periodic methods, outperforming the classical windowing methods.

- *Periodic excitation*: For periodic excitations, the LPM/LRM methodology allows the transients to be removed, so that there is no need to wait for the steady-state regime. This can result in a significant reduction of the measurement time [84].
- *Noise analysis*: High-quality noise power spectra estimates are obtained using the LPM/LRM if the input SNR is high. For measurements with comparable SNR at the input and output, it is also possible to make a full noise analysis resulting in high-quality (co)variance estimates (as a function of the frequency) of the input/output noise, provided that a good noise-free reference signal is available. If that is not the case, periodic excitations can be a solution to obtain good estimates (see the section "Stochastic Analysis of Periodic Excitations").
- *Leakage errors*: The LPM/LRM methods almost completely remove the leakage errors. For this reason, these methods are advised as the default choice for FRF measurements whenever random noise excitations are used. There is no reason to continue to use the classical windowing methods, except for very special cases where the memory of the processor would be very limited. The default setting to obtain a maximum leakage rejection and a minimal bias error is to choose the local bandwidth B in (37) as small as possible for a fixed degree (no user interaction is needed).
- *Disturbing noise sensitivity*: With the default settings (local bandwidth as small as possible), the noise sensitivity of the classical windowing methods and the LPM/LRM is comparable. For smooth systems, the local bandwidth can be increased without creating bias errors that are too large (this is an advanced form of smoothing over the neighboring frequencies). If needed, this gives an additional handle to the user to reduce the variance of the results [57].
- *Calculation time*: The calculation demands in real-time applications can be strongly reduced by completing the majority of the calculations before the measurements are done. An example of a real-time implementation on a microprocessor is discussed in [84].

MEASURING THE POWER SPECTRUM OF THE DISTURBING NOISE ON THE OUTPUT

The major problem in measuring the spectral properties of the disturbing noise is the separation of the true signals U_0, Y_0 and the noise N_U, N_Y. A first possibility is to use periodic signals, as discussed in the section "Stochastic Analysis of Periodic Excitations." A full nonparametric analysis of the (co)variance of the input and output noise could be made, without any interaction from the user [16], [47].

If the periodicity assumption does not hold, then an alternative approach is needed. In that case, the noise is separated from the signal using the estimated FRF, assuming that the input measurement is noise free. This can be generalized to noisy input and output measurements if an exactly known reference signal is available (for example, the signal in the memory of the generator) using the indirect method or the joint input–output method [85]–[87]. In that case, the SISO problem (from the input u to the output y) is replaced by a MIMO problem (from the reference r to the input u and the output y), and eventually it is still possible to obtain the (co) variance of the input-output noise. Motivated by the previous discussion, it is assumed in this section that the input is exactly known, while the output is disturbed by noise.

Assumption 5

No disturbing input noise

$$\begin{aligned} U(k) &= U_0(k), \\ Y(k) &= Y_0(k) + N_Y(k). \end{aligned} \tag{43}$$

Under these assumptions, only the variance $\sigma_Y^2(k)$ needs to be measured. This is done simultaneously with the FRF measurement. The classical windowing methods start from the measured coherence, while the new local parametric methods start from the residue analysis.

Coherence-Based Spectral Noise Analysis

Under Assumption 5, an estimate of the variance is obtained by [4]

$$\sigma_Y^2(k) = \hat{S}_{YY}(k) - \frac{|\hat{S}_{YU}(k)|^2}{\hat{S}_{UU}(k)} = \hat{S}_{YY}(k)(1-\gamma^2(k)). \tag{44}$$

The cross and auto spectra are estimated using the previous windowing methods, and the errors on the power spectrum estimate of the noise will be set by the leakage errors of $\hat{S}_{UU}$, $\hat{S}_{YY}$ [62], which are of $O(N^{-2})$ for the Hanning window.

Spectral Noise Analysis Using the Local Parametric Methods

The system description (S12) can be extended by a parametric noise model

$$Y(k) = G(k)U(k) + T_G(k) + N_Y(k),$$
$$= G(k)U(k) + T_G(k) + H(k)E(k) + T_H(k). \quad (45)$$

The output noise is modeled as a filtered white noise source with power spectrum $\sigma_e^2|H(k)|^2$ and a transient term $T_H(k)$ that models the leakage effects of the noise-shaping system H, completely similar to the role of $T_G(k)$ for the plant model. In practice, both transients $T_G(k)$, $T_H(k)$ are added together in the local parametric methods. The power spectrum $\sigma_e^2|H(k)|^2$ is directly estimated as the mean square value of the residuals in the local frequency band B (see the previous section) [19], [77], corrected for the degrees of freedom in the least square problem that define the ILRM, LPM, and the RLM. Since these residuals are no longer disturbed by the leakage effect, they have a much higher quality than those directly obtained using the coherence method (44). In this case, the errors on $\sigma_Y^2(k)$ are of the same order of magnitude as those of the corresponding local parametric method.

Spectral Noise Analysis: An Example

The output of a lowly damped, second-order system is disturbed by filtered white noise. To visualize all the important aspects, the noise filter was chosen with a resonance and an antiresonance. The results are shown in Figure 15 for the LPM and the coherence method using a Hanning window. Since the spectral resolution is quite high (many frequencies in the 3-dB bandwidth), the results with the LRM would be very similar. We note that the frequency resolution for the LPM method is 16 times higher than that of the coherence method. A smoother estimate can be obtained by averaging the results over neighboring frequencies in a sliding window.

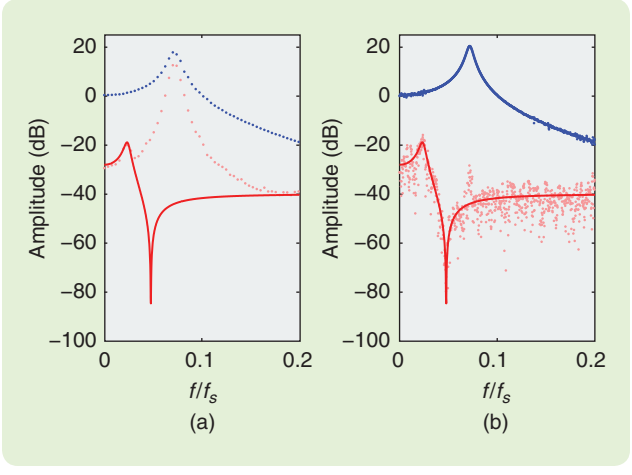


FIGURE 15 The estimation of the output noise variance σ_Y^2 using the coherence method with (a) a Hanning window and (b) the local polynomial method (LPM). A data record with a total length of 256×16 is split into 16 subrecords for the coherence method, while it is processed as one long record for the LPM approach. The true value of the power spectrum is plotted in red. The blue dots show the estimated frequency response function, and the pink dots show the estimated power spectrum. For the coherence method, the system dynamics are visible in the estimated power spectrum due to the poor separation of the system dynamics and the disturbing output noise. Moreover, the antiresonance is completely lost. This is not the case for the LPM method where the true noise spectrum is well followed by the estimates.

TIME-DOMAIN APPROACH

This section studies the direct measurement of the IR. In a second step, it is possible to obtain the FRF by calculating the FFT of the estimated IR. The frequency resolution can be increased by increasing the length of the time record using zero padding. The discussion herein assumes that the input measurement is not disturbed by noise (Assumption 5). The estimation of the IR function in the time domain is studied within a DT setting

$$y(t) = g(t) * u(t) = \sum g(k)u(t-k) = \sum g(t-k)u(k). \quad (46)$$

This restriction does not limit the generality of the discussion and the proposed methods, provided that either a ZOH input is used, or that the sampling frequency f_s is chosen high enough with respect to the bandwidth of the system, as was discussed in the section "Measurement Setup." Without loss of generality, the discussion is restricted to causal systems. The generalization to noncausal systems is straightforward.

Real-life systems have mostly an infinitely long IR that will be approximated by finite-length models $\tilde{g}(t), t = 0, 1, 2, \ldots, n$

$$\tilde{y}(t) = \sum_{k=0}^{n} \tilde{g}(k)u(t-k). \quad (47)$$

In contrast to the FRF models, the model length n needs to be specified. This seemingly increases the complexity of this approach (compared to the FRF methods), but in the latter case the frequency resolution f_0 needs to be well selected with respect to the system dynamics, as explained in "Link Between the Data Length and the Frequency Resolution of the Frequency Response Function Estimate," leading to a very similar problem.

In general, it is not possible to estimate the individual entries $g(t)$ in (47) independently from each other; all parameters are estimated at once. The IR estimation methods are global nonparametric methods. It is shown below that for some special choices of the input signal, the problem can be reduced to a local nonparametric problem, where each entry $\tilde{g}(t)$ is estimated individually.

IR Estimation, a Linear Least-Squares Problem

The underlying idea of all the methods discussed below is to solve the overdetermined set of equation [22]

$$y = K\tilde{g}, \tag{48}$$

with

$$y = (y(n), y(n+1), \ldots, y(N-1))^T, \tilde{g} = (g(0), g(1), \ldots, g(n))^T$$

and

$$K = \begin{pmatrix} u(n) & u(n-1) & u(n-2) & \ldots & u(0) \\ u(n+1) & u(n) & u(n-1) & \ldots & u(1) \\ u(n+2) & u(n+1) & u(n) & \ldots & u(2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ u(N-1) & u(N-2) & u(N-3) & \ldots & u(N-n-1) \end{pmatrix} \tag{49}$$

by minimizing the following least-squares cost function with respect to $g(k), k = 0, 1, 2, \ldots, n$

$$\begin{aligned} V &= \sum_{t=n}^{N} (y(t) - \tilde{y}(t))^2, \\ &= \sum_{t=n}^{N} \Big(y(t) - \sum_{k=0}^{n} \tilde{g}(t) u(t-k) \Big)^2. \end{aligned} \tag{50}$$

The sum (50) starts at $t = n$ to avoid unknown past values of the input that appear in (48). Alternatively, the unknown past values could be estimated too, but that would double the number of unknown parameters. Moreover, the product of the unknown parameters $\tilde{g}(k) \times u_{t<0})$ would turn the linear least-squares problem into a nonlinear one, and iterative methods would be needed to solve it. These choices are the time-domain equivalent of the leakage handling in the frequency-domain solutions.

No weighting function is added to the cost function, mainly because this would require prior noise information that is not available.

The minimizer of the cost function V (50) is obtained by solving [22]

$$K^T K\tilde{g} = K^T y, \tag{51}$$

with

$$K^T K = \begin{pmatrix} \hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \hat{R}_{uu}(2) & \ldots & \hat{R}_{uu}(n) \\ \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \ldots & \hat{R}_{uu}(n-1) \\ \hat{R}_{uu}(2) & \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \ldots & \hat{R}_{uu}(n-2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \hat{R}_{uu}(n) & \hat{R}_{uu}(n-1) & \hat{R}_{uu}(n-2) & \ldots & \hat{R}_{uu}(0) \end{pmatrix}, \tag{52}$$

and

$$K^T y = \begin{pmatrix} \hat{R}_{yu}(0) \\ \hat{R}_{yu}(1) \\ \hat{R}_{yu}(2) \\ \ldots \\ \hat{R}_{yu}(n) \end{pmatrix} \tag{53}$$

Link Between the Data Length and the Frequency Resolution of the Frequency Response Function Estimate

The frequency resolution δ_f of a frequency response function measurement is directly linked to the length of the subrecords N_{sub}

$$\delta_f = f_s/N_{sub} = Pf_s/N = P/T_{meas}, \tag{S21}$$

with P the number of subrecords, $N_{sub} = N/P$, f_s the sample frequency, and T_{meas} the total measurement time. The choice of P is a user choice set by the balance between the required frequency resolution and the total measurement time. A higher value of P allows for more averages and a lower standard deviation at the cost of a reduced frequency resolution for a given measurement time. (See Figure S12.)

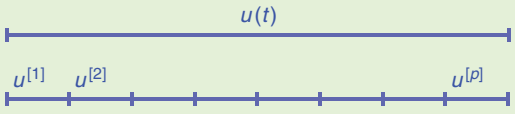


FIGURE S12 Splitting a long data record into P shorter subrecords reduces the original record length N to a subrecord length N/P. The corresponding frequency resolution drops from f_s/N to f_sP/N.

or

$$\hat{R}_{uu}\tilde{g} = \hat{R}_{yu}, \tag{54}$$

with

$$\begin{aligned} \hat{R}_{uu} &= \frac{1}{N-n} K^T K \in \mathbb{R}^{(n+1)\times(n+1)}, \\ \hat{R}_{yu} &= \frac{1}{N-n} K^T y \in \mathbb{R}^{(n+1)\times 1}. \end{aligned} \tag{55}$$

The size of the normal equations (54) is significantly smaller than the size of the original set (48) $K \in \mathbb{R}^{(N-n)\times(n+1)}$. Moreover, the matrices can be directly obtained from the sample estimates of the auto and cross-correlation, without any need to form the large matrix K, for example,

$$\hat{R}_{yu}(\tau) = \frac{1}{N-n} \sum_{k=n}^{N} y(k) u(k-\tau). \tag{56}$$

The solution of (54) will be discussed for different choices of the excitation signal. The latter choice strongly influences the complexity of the solution.

Remark 3

Instead of using delayed inputs, corresponding to z^{-k} as basis functions in the z-domain, it can be advantageous to use

more advanced basis functions. A first possibility is to use Laguerre functions [88]. It is possible to inject more knowledge (if the user has some prior idea about the pole locations of the system) using generalized orthogonal functions [89]. The major advantage of these methods is the reduction of the number of unknown parameters to be estimated, which results in a reduced uncertainty. Of course, it remains a major challenge to obtain valuable prior knowledge.

Impulse Response Estimation Using Impulse Excitations

For a long time, impulse (or step) excitations were the only practical possibility to create a wideband excitation. At the same time, (54) is significantly simplified. For a perfect impulse excitation,

$$u(t) = \sigma_u \delta(t), \tag{57}$$

and

$$R_{uu}(\tau) = \sigma_u^2 \delta(\tau). \tag{58}$$

$K^T K$ simplifies to

$$K^T K = \sigma_u^2 \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \tag{59}$$

and the normal equations (54) reduce to

$$\sigma_u^2 \tilde{g} = \hat{R}_{yu} = (y(0),\ y(1),\ y(2), \dots, y(n))^T, \tag{60}$$

which give direct access to the IR $\tilde{g}$ in the measured output. This approach was very popular prior to 1970 because no computers were needed. A paper recorder, combined with an impulse excitation, was enough to obtain a direct measurement of the IR.

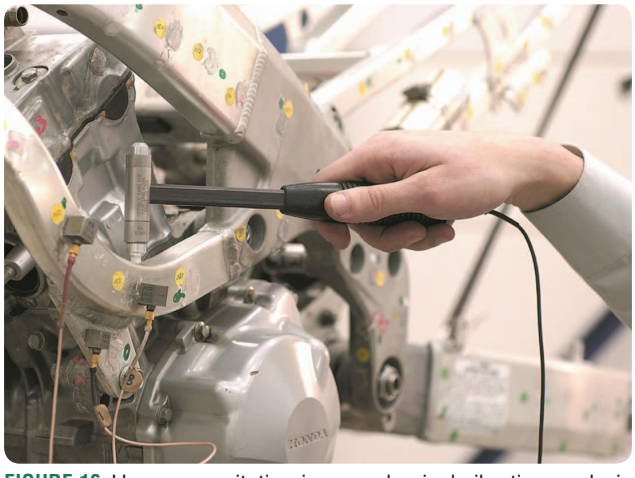


FIGURE 16 Hammer excitation in a mechanical vibration analysis in the vibration labs of Siemens Industry Software Leuven, Belgium. A force transducer is placed in the tip of the hammer that allows the impact force to be measured. Accelerometers are mounted on the mechanical structure to measure the response of the system. (Used with permission from Siemens.)

Modern Use of Impulse Excitations

Although the computational restrictions are completely removed today, the IR excitation is still popular in some application fields.

Mechanical Engineering

Mechanical engineers use special-instrumented hammer kits [18], [90]. The setup is struck by the hammer, and the IR is again directly measured (Figure 16). The major advantage (and the reason that it is still used) is that no complicated actuator connections need to be realized, and the excitation can be easily moved around the whole structure in a few minutes. The instrumented hammer kits also measure the actual applied force to compensate for variations in the excitation level. The duration of the pulse that sets the bandwidth of the excitation is controlled by a proper choice of the hammer tip [18].

Dynamic Calibration

During the development of primary standards for dynamic calibration problems, it is not possible to measure the input signal. Special setups have been built to create short impulses with controlled properties [91]–[94]. The main issue is to keep the duration of the impulse short enough, compared to the time constant of the sensor that needs to be calibrated.

Biomedicine

In the mathematical modeling of biomedical systems, the choice of input is usually very restricted, particularly when the subject is a human being. A single rapidly administered dose (administered either intravenously or orally) is the most commonly used form. It is then assumed that the duration of the administration is rapid compared with the kinetics of the subject, so that the input can be approximated by $Q\delta(t)$, where $\delta(t)$ is a delta function [95], [96]. For intravenous input, this form of input is usually referred to as a bolus dose. The only other form of input commonly used in such applications is a constant continuous infusion such as a step input [95], [96]. The unit step response is then the integral of the unit IR.

Averaging of Impulse Measurements

The direct IR measurement method works well only for measurements with high SNR because there is no averaging in the simple procedure. Increasing the impulse level is often impossible for practical reasons. Under low SNR measurement conditions, the results can be averaged over repeated experiments, but then it is extremely important to align the successive measurements using a good trigger signal or advanced signal processing methods to avoid smearing effects in the averaging procedure [97].

Remark 4

It sometimes may be easier to measure the step response, which is the integral of the IR [98].

Impulse Response Estimation Using White Noise Excitation

Not only do impulse excitations result in an impulse-like autocorrelation, but all wideband excitations with a flat amplitude spectrum have a similar behavior. For that reason, white noise excitations became a popular alternative for the impulse excitation.

Random White Noise Excitation

For simplicity, only white random noise is considered here. The actual distribution (for example, Gaussian noise or uniform noise) is not important from a theoretical point of view because the autocorrelation depends only on the second-order properties. In practice, the distribution will set the peak value of the excitation for a given power. It is very desirable to keep this value as low as possible to not overload the system and improve the SNR. In that case,

$$\lim_{N\to\infty} \hat{R}_{uu}(\tau) = E\{\hat{R}_{uu}(\tau)\} = \sigma_u^2 \delta(\tau), \tag{61}$$

and the same expressions as in the previous section are retrieved. However, the major difference with respect to the impulse excitation is that the system is persistently excited during the whole experiment. The variance on the estimate will drop as $O(1/N)$, so that even under low SNR conditions, good measurements can be obtained.

For finite-length records, the autocorrelation does not reach its expected value. It is a random variable of its own and eventually

$$\hat{R}_{uu}(\tau) = R_{uu}(\tau) + O\Big(\frac{1}{\sqrt{N}}\Big), \tag{62}$$

leading to errors in the estimated IR of the same order of magnitude. These can be eliminated by solving the original normal equations (54) at a loss of the simplicity. Nevertheless, it is still advantageous to use white noise excitations because these are the optimal excitation signals from the disturbing noise rejection point of view (see the section "Variance Analysis of the Frequency Response Function Measurements in the Presence of Disturbing Noise").

The cross-correlation is currently calculated using digital computers. These were previously unavailable, and analog correlators were built using tape recorders to delay the signals [99], [100] (see "Historical Note on Frequency Response Function and Impulse Response Measurements").

Well-Designed Deterministic Signals

The finite-length effects discussed in the previous section are further reduced by replacing the random excitation by a well-designed period excitation. Periodic signals have a periodic autocorrelation function. This creates no problem as long as the period length is longer than the useful length of the IR, for example five or ten times the dominating time constant of the system (see "Truncation Error of Finite Impulse Response Models").

A standard choice within this class of signals is well-designed binary signals [27], [28], [101], among others, the maximum length binary signals (see "Design of Excitation Signals") that could be generated using dedicated hardware, as shown in Figure S2. These signals combine many advantages:

- *Removal of random finite length effects*: The autocorrelation function is (see "Design of Excitation Signals")

$$\hat{R}_{uu}(\tau) = R_{uu}(\tau) + O\Big(\frac{1}{N}\Big). \tag{63}$$

  The error term $O(1/N)$ converges much faster to zero for an increasing data length N than the $O(1/\sqrt{N})$ term in (62) for white noise excitations. To fully retrieve these advantages, the clock frequency f_c should be 2.5 times larger than the highest frequency of interest.
- *Binary correlation*: Because the excitation signal is either 1 or –1, a binary correlator can be used. If only limited computer power is available (or in time-critical applications), this is a significant advantage.
- *Periodic signal*: All the advantages of periodic signals discussed in the section "Frequency Response Function Measurements Using Periodic Excitation" are also retrieved here, including simple averaging procedures and direct access to the noise power spectrum.

A disadvantage of the maximum length binary sequence (MLBS) (and the related sequences) is the high sensitivity to nonlinear distortions compared to randomized excitations like random noise or random phase multisines [16]. These create large spikes in the estimated IR [102]. Averaging over multiple realizations of the MLBS in combination with a median filter can largely remove these spikes [102], but this requires much longer measurement times. Another alternative is to use randomized periodic ternary signals that do not excite the multiples of the second and third harmonic [103].

The advantages of correlation methods combined with PRBS excitations was previously recognized [14] and applied to industrial processes like gas chromatography [26], nuclear power plants [104], and oil refineries [105].

Impulse Response Estimation Using Arbitrary Excitations

The direct IR estimation method (54) can be combined with arbitrary excitations. Calculating the solution of (54) can be sped up by exploiting the underlying Toeplitz nature of the matrices [106].

Historical Note on Frequency Response Function and Impulse Response Measurements

Before the 1950s and 1960s, it was not possible to estimate the frequency response function (FRF) or impulse response (IR) from experimental data obtained during the normal operation of a system. Specific experiments were made to measure directly the IR or the FRF of the system. The IR was obtained using an impulse, while sine excitations were used to measure the FRF at a given frequency [37]. The main reason to use these simple methods was the lack of analog-to-digital converters, so that all calculations needed to be done with analog circuits or computers. The latter were very tedious to use, due to saturation and drift problems. In a more advanced setting for mechanical vibration measurements, a swept sine excitation was used in combination with a tracking filter tuned at the instantaneous sine frequency. The filter reduces (nonlinear) disturbances and noise [18]. In the late 1950s and the start of the 1960s, analog correlators were built to measure the IR using wideband random excitations [99], generalizing the possibilities of sine correlators [100]. These results were significantly improved by using pseudorandom binary sequence (PRBS) generators (see Figure S2). These were simple to build, including the generation of arbitrarily delayed copies, using the electronic components that were available at that time. In combination with the analog correlators, it became possible to obtain strongly improved IR estimates in a shorter time.

A shift toward digital signal processing was eventually made, and the analog correlators were replaced by digital signal processing algorithms. Much time and effort was spent educating the engineering community to understand the required signal theory, in particular correlation methods. Although the Fourier transform was known among radio engineers [39], it was not until the 1960s that it became a practical engineering tool. The fast Fourier transform (FFT) was first published in 1965 [60] (see also [41] and [42]), the same time that digital computers became widely available. This was the start of a new era in FRF measurements, using spectral estimation methods that were efficiently implemented to minimize the size of the required core memory of the computer [1]. The cost to calculate a 2^{10}-point Fourier transform dropped from a few dollars to hundredths of a cent, while the calculation time was reduced from several minutes to tens of milliseconds [42]. This was the start of digital signal processing. These methods became the dominant approach in most fields, and they are still the standard tools that are implemented in today's commercial equipment [3], [4], [18]. A renewed interest in developing more advanced FRF and IR has recently been experienced. These new activities make maximum use of the greatly increased computing power, which allowed for a shift toward methods that no longer suffer from leakage and are less sensitive to disturbing noise.

To give a better idea of the problems that were faced at that time, the measurement of the FRF using a sine excitation and the measurement of the IR using PRBS signals and analog correlators are discussed in more detail.

MEASURING THE FREQUENCY RESPONSE FUNCTION USING SINE EXCITATIONS

This section is based on the methods reported in [37] and [40]. Detailed information on the historical methods can be found in [40]. Under steady-state conditions, the output of a linear dynamic system excited by $u(t) = A\sin(2\pi ft + \phi)$ is

$$zy(t) = A\,|G(f)|\sin(2\pi ft + \phi + \angle G(f)). \tag{S22}$$

Graphical Method

The input and output are recorded on a plotter. Using visual inspection, the amplitude $|G(f)|$ is obtained as the ratio of the measured amplitudes, while the phase is estimated from the phase difference of both signals. These methods can be refined using Lissajous plots [37]. For poor signal-to-noise ratio (SNR), the results become very noise sensitive. For that reason, correlation methods were developed that allow the signals to be averaged over time to improve the SNR.

Variance Analysis of the Direct Impulse Response Method

The covariance matrix C_g is directly available for the linear least-squares problem

$$C_g = (K^TK)^{-1}K^TC_{n_y}K(K^TK)^{-1}. \tag{64}$$

Asymptotic Frequency-Domain Expression

An asymptotic frequency-domain interpretation for the model order $n \to \infty$ of this result can be made [21]

$$\sigma^2_{G_{IR}}(k) \approx \frac{n}{N}\frac{\sigma^2_Y(k)}{S_{UU}(k)}. \tag{65}$$

The subscript IR indicates that this result is obtained via the IR estimate. Equation (65) shows that the length of the estimated IR should be well balanced: as small as possible to reduce the variance, but long enough to keep the systematic errors (bias) small. From (65), it follows also that the standard deviation $\sigma_{G_{IR}}$ drops again as $O(1/\sqrt{N})$.

Remark 5

More general variance expressions for (65) exist that are valid for finite-model orders [107]. The simple expression (65) will be further used because it will lead later to a very user-friendly rule of thumb.

Correlation Method

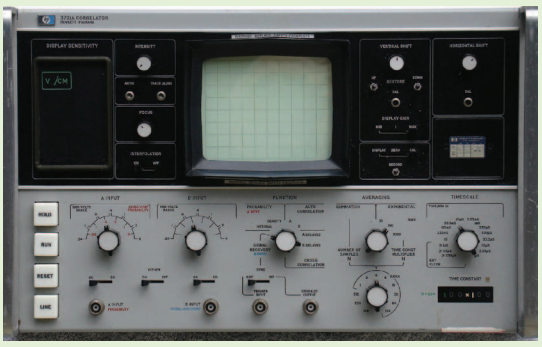
Two analog correlation methods were used to measure the FRF at a given frequency. The first correlation technique, called *quadrature components method* [37], makes a double correlation with the sin and cos. This implementation typically requires a correlation time equal to five times the dominant time constant of the system [40]. The second correlation method, called the *phase null method*, aligns the correlation signal first with the output signal (phase null) by manually tuning a potentiometer meter, then beginning a two-step correlation. This method requires more time but results in more precise measurements.

Today, sine-based network analyzers are still popular in the microwave field, partly because it is still very expensive to generate arbitrary signals at very high frequencies (today, this is above 20–30 GHz), and the SNR is much higher because narrow band filters can be applied.

MEASURING THE IMPULSE RESPONSE USING PSEUDORANDOM BINARY SEQUENCE SIGNALS AND DEDICATED GENERATION/CORRELATION HARDWARE

Measuring the FRF using sine excitations is very time consuming, especially at low frequencies and for systems with long time constants. Only one frequency is probed at a time, and after each frequency change, the system should reach steady-state conditions before the next measurement can be made. Using white noise excitations, all dynamics of the system are excited at once, but correlation methods are needed to retrieve the IR from these measurements. In practice, the random white noise excitation is replaced by a PRBS signal (see "Design of Excitation Signals") that has a deterministic and almost flat amplitude spectrum. Special hardware was developed [13], and combinations of PRBS generators and analog correlators (as shown in Figure S13) were commercialized [38], [99]. A few years later, when FFT's could be implemented on microprocessors, dedicated displays were built that could show the FRF on their screen in amplitude and phase [43]. This became the first generation of dynamic signal analyzers, which were replaced in the mid-1970s by single-box instruments that were based on spectral analysis methods (see the section "FRF Measurements Using Random Excitations") [58], [59].

(a)



(b)

FIGURE S13 The combination of a (a) pseudorandom binary signal generator and (b) correlator from the 1960s. Besides the original signal $u(t)$, a copy with a user adjustable delay $u(t-\tau)$ was also generated. This allowed the correlation $y(t)u(t-\tau)$ to be measured using the analog correlator in (b), giving a direct measurement of the impulse response. In the early 1970s, the fast Fourier transform of the impulse response was calculated and shown on a "Spectrum Display" The combination of the generator, correlator, and spectral display resulted in the first dedicated frequency response function measurement setups [38], [43], [99]. (Courtesy of Kenneth Kuhn, http://www.hpmemoryproject.org.)

Comparison with the Spectral Analysis Methods

In (28), the variance on the FRF $G(k)$ obtained with the spectral methods is given. In the absence of input noise ($\sigma_U = 0, \sigma_{YU} = 0$), this expression simplifies to

$$\sigma^2_{G_{FRF}}(k) = \frac{1}{P}\frac{\sigma^2_Y(k)}{S_{UU}(k)}. \tag{66}$$

The subscript FRF indicates that this result is obtained via a direct FRF measurement. The length of the subrecords is $N_{\text{sub}} = N/P$. The additional gain for overlapping subrecords (up to a factor two) can be added to (66). The additional loss [up to a factor of 4^2 (see Figure 13)] is not considered for simplicity, and the leakage-induced variance is also disregarded.

From (65) and (66), it follows that

$$\frac{\sigma^2_{G_{FRF}}(k)}{\sigma^2_{G_{IR}}(k)} = \frac{1}{2}\frac{\frac{1}{P}}{\frac{n}{N}} = \frac{N_{\text{sub}}}{2n}, \tag{67}$$

with $N_{\text{sub}} = N/P$ the length of a subrecord in the spectral analysis method and n the length of the estimated IR.

This result is very revealing because it compares the FRF and the IR-based methods, starting from their experiment design parameters. The choice of N_{sub} and n is directly

Truncation Error of Finite Impulse Response Models

Most physical stable systems have an impulse response (IR) that decays exponentially to zero. In theory, such an impulse response is infinitely long. In practice, it can be arbitrarily well approximated by a finite impulse response (FIR) model. This is very attractive from a modeling point of view. The length of the FIR that is needed to keep the truncation error below a given level can be normalized on the time constant of the system. For a first-order system with an IR $g(t) = \alpha e^{-t/\tau}$ excited by white noise and truncated at time T, the relative root mean square truncation error is $e_{\text{RMS}} = e^{-T/\tau}$. This error is shown in Figure S14 as a function of the truncation time. Truncating the impulse response at T equal to $2\tau, 5\tau,$ and 7τ results respectively in a relative RMS error below 10, 1, and 0.1%.

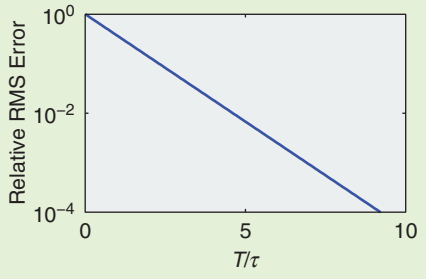


FIGURE S14 The relative root mean square (RMS) error induced by truncating the impulse response of a first-order system with time constant τ.

linked to practical user choices. Using the results in "Link Between the Data Length and the Frequency Resolution of the Frequency Response Function Estimate," the following remarks can be made.

Choice of n

The length of the estimated IR should be long enough such that the truncation bias is small, as discussed in "Truncation Error of Finite Impulse Response Models." The relative mean-squared error (MSE) due to the truncation-induced model errors is $e^{-2n/\tau}$. Truncating the IR at n equals $2\tau, 5\tau, 7\tau$ results, respectively, in an RMS error below 10%, 1%, and 0.1%.

Selecting the Subrecord Length N_{sub}:

The frequency resolution of the spectral method is $\triangle_f = 1/N_{\text{sub}}$. In "Characterizing a Resonance by Its 3-dB Bandwidth," it is explained that the required resolution depends on the 3-dB bandwidth of the system. This is $B_{3\text{dB}} = 1/\pi\tau$ for a resonating system. Hence, the relative frequency resolution of G_{FRF} is $\triangle_f / B_{3\text{dB}} = \tau/2N_{\text{sub}}$, and the number of measured frequencies in the 3-dB band is

$$F_{3\text{dB}} = \frac{N_{\text{sub}}}{\pi\tau}. \tag{68}$$

Substituting both results in (67), shows that for the 1% error level ($n = 5\tau$), the following ratios hold

$$\frac{\sigma^2_{G_{FRF}}(k)}{\sigma^2_{G_{IR}}(k)} = \frac{\pi\tau F_{3\text{dB}}}{10\tau} \approx \frac{F_{3\text{dB}}}{4}. \tag{69}$$

It is concluded that for a very low frequency resolution (four frequencies in the 3-dB bandwidth) and 1% RMS bias error, both the G_{FRF} and G_{IR} require about the same experiment length. For more noisy data (higher errors can be tolerated, and so n can be reduced to 2τ) or for higher-frequency resolutions, the IR approach becomes more attractive if the calculation time and the size of the computer memory are not an issue.

Variance Reduction by Combining Data and Prior Knowledge: Regularization

The linear least-squares estimate (54) is the maximum likelihood estimate for white Gaussian disturbing noise n_y. Under these conditions (and noting the problem is linear-in-the-parameters), it is well known that this estimator reaches the Cramer–Rao lower bound [16], [21], [22], even for finite sample lengths. It is impossible to find another unbiased estimator that will have a smaller uncertainty (covariance matrix) [16], [21], [22]. However, it is also well known that the variance can be further reduced by allowing a (small) bias [15] (see "Bias and Variance Tradeoff of Estimators"). This observation is at the start of the regularization methods [15]. The cost function (50) is extended with an additional regularization term $\gamma \tilde{g}^T P^{-1} \tilde{g}$

$$\begin{aligned} V &= \sum_{t=n}^{N} (y(t) - \tilde{y}(t))^2 + \gamma \tilde{g}^T P^{-1} \tilde{g}, \\ &= \sum_{t=n}^{N} \left(y(t) - \sum_{k=0}^{n} \tilde{g}(k) u(t-k) \right)^2 + \gamma \tilde{g}^T P^{-1} \tilde{g}. \end{aligned} \tag{70}$$

The regularization term pulls the estimates toward zero, resulting in a reduced output variance at a cost of an increasing bias b. For some value of γ, the mean squared error $|b|^2 + \sigma^2$ will reach a minimum, as shown in Figure S7. From practical experience, a large variance reduction can be obtained for a small increase in the bias.

The success of this approach strongly depends on a proper choice of the regularization matrix P. This choice can be guided by putting the regularized cost function (70) in a Bayesian framework. A comprehensive introduction to this approach is given in [15]. The Bayesian framework allows the regularization matrix P to be connected to the prior knowledge of the user. For the IR estimation problem, very valuable prior information is stability and smoothness, both of which are discussed below.

Frequency Response Function Measurements for MIMO Systems

All the results for single-input, single-output (SISO) frequency response function (FRF) measurements also hold for the measurement of the multiple-input, multiple-output (MIMO) frequency response matrix (FRM) at each frequency k

$$\boldsymbol{G}(k) \in \mathbb{C}^{n_y \times n_u}, \tag{S23}$$

with n_u and n_y the number of inputs and output of the system. The major change with respect to the SISO setup is the need to separate the effect of the multiple inputs on a given output. The number of outputs does not add to the complexity of the problem. Consider, without loss of generality, a system with $n_u = 3$ inputs and $n_y = 1$ output

$$Y(k) = (G_1(k) \;\; G_2(k) \;\; G_3(k)) \begin{pmatrix} U_1(k) \\ U_2(k) \\ U_3(k) \end{pmatrix}. \tag{S24}$$

The FRM (S24) cannot be identified from one experiment unless constraints are imposed on either the excitation signals U_1, U_2, U_3 or the smoothness of the system. For uncorrelated inputs, it is possible to consider two unmeasured inputs as a disturbance, but this will increase the uncertainty [4]. Alternatively, three or more experiments can be combined to resolve the problem. Each of these options is briefly discussed below.

FREQUENCY RESPONSE MATRIX MEASUREMENTS USING A SINGLE EXPERIMENT

Zippered Multisines

A first possibility to separate the different input channels would be to excite one input at a time. This results in a drop of the signal-to-noise ratio because the other inputs are set to zero. An alternative implementation of this idea is to realize the separation in the frequency domain by assigning each excitation frequency to only one input. This leads to zippered multisines. The pth input u_p is exciting only the frequencies $k = p + n * n_u$, with $n \in \mathbb{N}$. This choice decouples the inputs in (S24) over the frequency, and at the selected frequencies the SISO methods can be applied. The cost is a reduction of the frequency resolution with a factor n_u. The method also requires complete control over the excitation signal; input distortions that disturb the zeros in the spectra also destroy the decoupling properties [5].

Local Parametric Methods

Using the local parametric methods (generalized to MIMO systems) is an alternative to separate the different inputs [20], [77]. If the input is sufficiently rich (for example, random phases), the local parameterized models can still be identified from a single experiment by combining the neighboring frequencies in one set of equations. This is a very attractive and robust alternative to the other approaches that are discussed in this sidebar. The properties of this approach remain the same as those discussed in the section "Improved Frequency Response Function Measurements Using Local Parametric Methods." A detailed analysis is made in [20]. The more inputs that are considered, the larger the local bandwidth B that is used in the local parametric methods.

FREQUENCY RESPONSE MATRIX MEASUREMENTS USING MULTIPLE EXPERIMENTS

In this approach, $n_e \geq n_u$ experiments are made, and the spectra of the input and output at a given frequency k are stored as the columns of the matrices $\boldsymbol{U} \in \mathbb{C}^{n_u \times n_e}$ and $\boldsymbol{Y} \in \mathbb{C}^{n_y \times n_u}$. The FRM at frequency k is then retrieved as the least-squares solution [6]

$$\hat{\boldsymbol{G}} = \boldsymbol{Y}\boldsymbol{U}^H(\boldsymbol{U}^H\boldsymbol{U})^{-1}. \tag{S25}$$

The condition number $\kappa_{\boldsymbol{U}}$ of $(\boldsymbol{U}^H\boldsymbol{U})$ has a direct impact on the covariance matrix of $\hat{\boldsymbol{G}}$, which grows rapidly with the number of inputs n_u [6]. For (white) noise excitations, the conditioning can be improved by collecting more experiments n_e, but this comes at a cost of an increased measurement time. For that reason, alternative experimental procedures were proposed. The first group works for $n_u = 2^n$ and is based on the use of repeated inputs with well-selected sign switches that are generated using Hadamard matrices [7], [8]. These results were further generalized to an arbitrary number of inputs [6]. Alternatively, well-designed binary signals can be used [9]. All of these methods result in a condition number $\kappa_{\boldsymbol{U}} = 1$, which leads to very strong reductions of the uncertainty for systems with a large number of inputs. For example, the standard deviation is reduced by a factor of ten for a system with ten inputs.

- *Stability*: If it is known that the unknown system is stable, it follows that the impulse response should decay exponentially to zero for lumped systems [108].
- *Smoothness*: Most IRs of physical systems have an intrinsic smoothness. This can be expressed by requiring that neighboring points are correlated to each other.

Both aspects can be translated in a proper choice of the regularization matrix P. A standard choice is the tuned or correlated kernel

$$P_{k,j} = \lambda \alpha^{\max(k,j)}. \tag{71}$$

The hyperparameters α, λ are tuned on the data [15]. In [109], it is shown that this regularization can be interpreted

as a restriction on the variability of the low-pass filtered IR that is forced exponentially to zero.

This simple idea is very powerful. The additional information that is added by the prior knowledge of the data is very significant, and it results in a strong reduction of the uncertainty of the simulated system output at a cost of a (small) bias on the estimated IR. Overall, the MSE on the output is minimized by these methods. However, this bias is focused at the (dominant) resonance, as shown in Figure 17, leading to an underestimation of the damping at those resonances. This is unacceptable in applications where a precise damping estimate is an important design input. This conflicting observation is mainly due to the different objective of regularization methods (the objective is to minimize the MSE of the modeled system output) and FRF measurements (the objective is a reliable FRF measurement at all frequencies of interest).

Comparison of the Hanning, Least-Squares, and Regularized Least-Squares Method

This section estimates the FRF using the Hanning method with an overlap of 2/3, and the results are compared with the FRF obtained from the estimated IR using the least-squares and the regularized least-squares.

Simulation 1: A Resonating System

The system is excited with filtered white noise up to $0.4\,f_s$. The output is disturbed with white noise so that the global SNR is 6 dB (noise power at 25% of the signal power at the output). The RMS error is obtained from 200 repeated simulations. The results are shown and discussed in Figure 17. From these results, it follows that the nonregularized least-squares (LS) gives the best in-band results if no (local) bias is tolerated around the dominant resonance frequency. If that is no problem, the regularized least-squares is the best. The errors grow very fast at high frequencies because the excitation power drops to zero, resulting in a very low SNR of the measurements. In this band, the regularized method reduces the error by 20 dB or more.

Simulation 2: A First-Order System with Delay

The system is excited with filtered white noise up to $0.4\,f_s$. The output is disturbed with white noise so that the SNR in the passband of the system is 10 dB. The RMS error is obtained from 1000 repeated simulations. The results are shown and discussed in Figure 18. From these results, it follows that the regularized LS gives the best results, followed by G_{LS} with the IR length fixed to $n = 8\tau = 40$. Tuning the IR length by minimizing the error on a validation set increased the in-band error with 2 dB. This shows again that, from the FRF point of view, it is not obvious how to select the optimal IR length. The errors of the Hanning method are ten times larger while its spectral resolution is reduced by a factor of eight, because the original record was split in eight subrecords. This shows that, for systems with slowly varying dynamics, a huge gain can be made by replacing the classical methods with the recently developed tools. In this case, the spectral methods would need 100 times more measurements to realize a similar quality as the (regularized) LS methods.

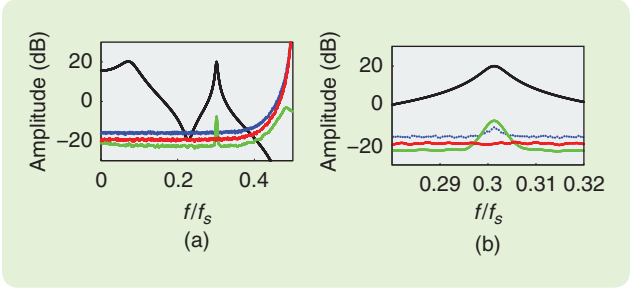


FIGURE 17 A comparison of three frequency response function (FRF)-measuring methods. The true FRF (black), and the observed root mean square (RMS) error ($\sqrt{\text{Bias}^2 + \text{variance}}$) of the Hanning (blue), IR_{LS} (red), and regularized IR_{reg} (green) methods are shown. The true system consists of the sum of two resonating systems having respectively a time constant $\tau_1 = 7$ and $\tau_2 = 66$. The Hanning method uses subrecord lengths of $N_{\text{sub}} = 2048$, and the total data length is $16 \times N_{\text{sub}}$. An overlap of 2/3 is used, resulting in an additional reduction of the variance with a factor two. For the IR_{LS}, the impulse response length was set to $n = 7 \times \tau_2 = 462$. Overall, the RMS error of IR_{reg} is the smallest, especially at large frequencies where the signal-to-noise ratio becomes very low due to the filtered input. This comes with a large bias around the second (dominant) resonance [see also the zoom in (b)]. The RMS error of the Hanning method is the largest, and this method also has a significant bias around the second resonance. The impulse response method shows no bias and has a smaller RMS error than the Hanning method.

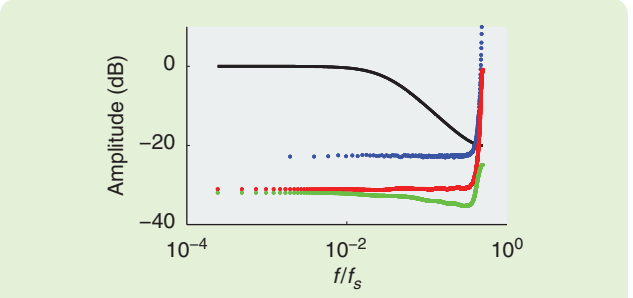


FIGURE 18 A comparison of three frequency response function (FRF)-measuring methods. The true FRF (black), and the observed root mean square (RMS) error ($\sqrt{\text{Bias}^2 + \text{variance}}$) of the Hanning (blue), IR_{LS} (red), and regularized IR_{reg} (green) methods are shown. The true system mimics a process consisting of a first-order system with a time constant $\tau = 5$ and a delay of 2τ. The Hanning method uses subrecord lengths of $N_{\text{sub}} = 512$, and the total data length is $8 \times N_{\text{sub}}$. An overlap of 2/3 is used, resulting in an additional reduction of the variance with a factor of two. For the IR_{LS} and IR_{reg}, the impulse response length was set to $n = 8 \times \tau_2 = 40$. It can be seen that, overall, the RMS error of IR_{reg} is the smallest with a gain of 1 dB with respect to IR_{LS} and approximately 20 dB with respect to the Hanning method. At the high frequencies (where the signal-to-noise ratio is very low due to filtered input), the gain is even larger. The frequency resolution is eight times higher for the least-squares methods than the Hanning method.

The classical spectral analysis methods still dominate the field, although it is possible perform significantly better with the recently developed local parametric methods or direct IR estimation methods.

Remark 6

The method is not changed to account for the delay. It will be accommodated automatically in the IR estimate: $\hat{g}(k) \approx 0$, $\hat{g}(k) \approx 0, k \leq delay$ at the start of the estimated IR.

User Guidelines

- *Direct estimation of the IR*: This method leads to high-quality estimates with a high noise rejection and low bias errors, provided that the selected IR length is long enough.
- *Selection of the IR length*: 1) For systems with fast-varying dynamics like resonating systems, classical model complexity selection methods Akaike's information criterion and Bayesian information criterion [21], [22] cannot be used for the purpose of FRF estimation because these methods balance variance versus bias. This bias is concentrated around the dominant resonance frequencies, leading to an over estimation of the damping. 2) For systems with slowly varying dynamics, the regularized LS methods can further improve the results of the LS method.
- *Tuning the bias error*: The user can set the lowest damping level that should be estimated without bias. On the basis of that result, the length of the IR $n = \alpha\tau$ in the least-squares method can be set; see "Truncation Error of Finite Impulse Response Models."
- *Regularization*: In the frequency bands with a low SNR, the results can be significantly improved using the regularization approach. This may bias those frequencies where the dynamics vary rapidly.

PUBLICLY AVAILABLE SOFTWARE

Most of the results can be reproduced using (free) public software.

- The book [73] provides an exercise-based introduction to system identification, dealing, among others, with the generation and analysis of excitation signals (Chapter 2) and FRF measurements (Chapter 3). The Matlab solutions of the exercises are available on the book support site booksupport.wiley.com. The software to generate the following figures (or similar figures) can be downloaded from this website: Figure 7 and 8 (Exercise 37), Figure 9 (Exercise 29), Figure 11 (Exercise 30), and Figure 14 and 15 (Exercises 51 and 52).
- The results in Figure 17 and 18 on IR estimation were obtained using the Matlab system identification toolbox, using the routine "arxRegul." This toolbox has also a routine "etfe" to calculate the ETFE.
- The freely available frequency domain identification toolbox FDIDENT can be used to generate the advanced excitation signals that are discussed in this article (Figure S5), and to create the nonparametric noise analysis using periodic excitations (Figure 6)
- Code to generate special (binary and multi-level) signals is available at https://www2.warwick.ac.uk/fac/sci/eng/research/systems/bbsl/signal_design [28].

CONCLUSION

This article provides a tutorial overview of IR and FRF estimation methods. The historical approaches are discussed, together with recently developed methods. It is shown that the available computer resources strongly influence the choice of the best-suited algorithm to solve the problem. The classical spectral analysis methods still dominate the field, although it is possible to perform significantly better with the recently developed local parametric methods or direct IR estimation methods. Using these methods, shorter experiments can be used to obtain a similar or even better quality of the results.

The discussion of the spectral methods is not the same as the classical textbooks. The authors preferred to provide a new interpretation of the properties of these methods, starting from the recent insights in the structured nature of leakage errors.

A short historical overview is included, showing that IR and FRF measurements have been completed for a long time. There was a strong evolution driven by the major changes in the available technology, starting with graphical methods and including analog correlation methods, to be eventually driven by digital signal processing methods. Too many engineers and scientists are not aware of the new possibilities that are available today, which leads to a waste of efforts and resources. This article aims to fill this gap.

A final set of user guidelines is formulated, directing the user to the best choice based on the available computing facilities.

- *Store the reference signal together with the data*: Whenever the external reference signal is available, it is strongly advised to store it together with the data. This will be very useful whenever there are closed loops in the system to be tested, including interactions between the generator and the setup.
- *Use periodic excitations whenever it is possible*: Periodic excitations give access to a full nonparametric noise model, even under closed-loop experimental conditions.

Combination with the local parametric methods or the IR method allows the user to eliminate transient effects if needed (high computational cost). Otherwise, a direct division of the output/input spectra can be made (lowest computational cost).

- *Use the direct IR estimation for systems with rapidly varying dynamics and low SNR measurements*: The direct IR method gives the best noise and leakage rejection. Care should be taken that the IR length n is well tuned (such as five to ten dominant time constants of the system, depending upon the required precision). The calculation effort grows with n as $O(n^3)$ without special numerical implementations.
- *Use the regularized IR estimation for systems with slowly varying dynamics and low SNR measurements*: The regularized IR reduces the mean square error of the direct IR method. Moreover, this method protects intrinsically against overmodeling, so that the choice of the IR length becomes less critical.
- *Use the local parametric methods for high SNR measurements*: If the SNR of the data is (very) high, the local parametric methods are a very attractive alternative for the direct IR methods. High-quality results are produced without needing to solve large sets of equations.
- *Real-time applications*: The local polynomial method is well suited for real-time applications. For a fixed input signal, all the matrix inverses can be precalculated and stored.
- *Windowing methods*: These methods require the least computer resources of all the methods. For FRF measurements, the combination of a sine window with a 66% overlap is the best combination. These methods have the highest sensitivity to leakage errors. It is advised to use these methods as a last resort if the calculation time is very critical and the LPM methods are still too demanding.

ACKNOWLEDGMENTS

The authors acknowledge Bart Peeters from LMS International, part of Siemens Product Lifecycle Management, for the photo in Figure 16. This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), the Flemish Government (Methusalem), the Vrije Universiteit Brussel (VUB), and the ERC advanced grant SNLSID, under contract 320378.

AUTHOR INFORMATION

Johan Schoukens (Johan.Schoukens@vub.be) received the master's degree in electrical engineering in 1980 and the Ph.D. degree in engineering sciences in 1985, both from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. In 1991, he received the degree of Geaggregeerde voor het Hoger Onderwijs from VUB, and in 2014 the doctor of science degree from the University of Warwick. From 1981 to 2000, he was a researcher for the Belgian National Fund for Scientific Research in the VUB Electrical Engineering Department. From 2000 to 2018, he was a full-time professor in electrical engineering, and since 2018 he has served as an professor emeritus in the VUB INDI Department. In 2018, he also became a member of the Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands. From 2009 to 2016, he was visiting professor at the Katholieke Universiteit Leuven, Belgium. His main research interests include system identification, signal processing, and measurement techniques. He is a Fellow of the IEEE. He received the 2002 Andrew R. Chi Best Paper Award of the *IEEE Transactions on Instrumentation and Measurement*, the 2002 Society Distinguished Service Award from the IEEE Instrumentation and Measurement Society, and the 2007 Belgian Francqui Chair at the Université Libre de Bruxelles, Belgium. He is a member of Royal Flemish Academy of Belgium for Sciences and the Arts. In 2011, he received a doctor honoris causa degree from the Budapest University of Technology and Economics, Hungary. Since 2013, he has been an honorary professor of the University of Warwick.

Keith Godfrey was appointed to the academic staff of the School of Engineering in 1973, and he is now a professor emeritus in the School of Engineering at the University of Warwick. He received the doctor of science degree from the University of Warwick in 1990 for publications with the collective title "Applications of Modeling, Identification and Parameter Estimation in Engineering and Biomedicine." He is the author of a book on compartmental modeling published by Academic Press in 1983 and is the (co)author of more than 200 papers. He is a member of the International Federation of Automatic Control Technical Committees on Biomedical Engineering and Control, and on Modeling, Identification and Signal Processing. He is currently working on the book *Industrial Process Identification: Perturbation Signal Design and Applications*, coauthored with Ai Hui Tan (Multimedia University, Malaysia), to be published by Springer in their book series Advances in Industrial Control.

Maarten Schoukens received the master's degree in electrical engineering: electronics and information technology and the Ph.D. degree from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 2010 and 2015, respectively. From 2015 to 2017, he was a postdoctoral researcher with the Electrical Engineering Department, VUB. In October 2017, he joined the Control Systems research group, TU/e, Eindhoven, The Netherlands, where he is currently working as a Marie Skłodowska-Curie Individual Fellow. His interests include the measurement and identification of linear parameter-varying and nonlinear systems.

REFERENCES

[1] P. D. Welch, "A direct digital method of power spectrum estimation," *IBM J. Res. Develop.*, vol. 5, no. 2, pp. 141–156, 1961.

[2] P. D. Welch, "The use of fast Fourier transform for estimation of power spectra: A method based on time averaging over short modified periodograms," *IEEE Trans. Audio Electroacoust.*, no. 2, pp. 70–73, 1967.

[3] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*. New York: Wiley, 1971.

[4] J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlations and Spectral Analysis*. New York: Wiley, 1980.
[5] D. E. Rivera, H. Lee, H. D. Mittelmann, and M. W. Braun, "Constrained multisine input signals for plant-friendly identification of chemical process systems," *J. Process. Control*, vol. 19, no. 4, pp. 623–635, 2009.
[6] T. Dobrowiecki, J. Schoukens, and P. Guillaume, "Optimized excitation signals for MIMO frequency response function measurements," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, pp. 2072–2079, 2006.
[7] P. A. N. Briggs and K. R. Godfrey, "Pseudorandom signals for dynamic analysis of multivariable systems," *Proc. Inst. Elec. Eng.*, vol. 113, no. 7, pp. 1259–1267, 1966.
[8] P. Guillaume, R. Pintelon, and J. Schoukens. "Accurate estimation of multivariable frequency response functions," in *Proc. Int. Federation Automatic Control World Conf.*, San Francisco, CA, July 1–5, 1996, vol. I, pp. 423–428.
[9] A. H. Tan, H. A. Barker, and K. R. Godfrey, "Identification of multi-input systems using simultaneous perturbation by pseudorandom input signals," *IET Control Theory Appl.*, vol. 9, no. 15, pp. 2283–2292, 2015.
[10] R. Voorhoeve, A. van der Maas, and T. Oomen, "Non-parametric identification of multivariable systems: A local rational modeling approach with application to a vibration isolation benchmark," *Mech. Syst. Signal Process.*, to be published.
[11] P. E. Wellstead, "Non-parametric methods of system identification," *Automatica*, vol. 17, no. 1, pp. 55–69, 1981.
[12] K. R. Godfrey, "Correlation methods," *Automatica*, vol. 16, no. 5, pp. 527–534, 1980.
[13] A. D. G. Hazlerigg and A. R. M. Noton, "Application of cross-correlating equipment to linear-system identification," *Proc. Inst. Elec. Eng.*, vol. 112, pp. 2385–2400, Dec. 1965.
[14] K. R. Godfrey, "Theory and application of pseudo-random sequences," *Control*, no. 6, pp. 305–308, June 1966.
[15] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.
[16] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. Hoboken, NJ: Wiley, 2012.
[17] P. E. Wellstead, "Reference signals for closed-loop identification," *Int. J. Control*, vol. 26, no. 6, pp. 945–962, 1977.
[18] D. J. Ewins, *Modal Testing: Theory and Practice*. New York: Wiley, 1991.
[19] J. Schoukens, G. Vandersteen, K. Barbe, and R. Pintelon, "Nonparametric preprocessing in system identification: A powerful tool," *Eur. J. Control*, vol. 15, pp. 260–274, May-Aug. 2009.
[20] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe, "Estimation of nonparametric noise and FRF models for multivariable systems. Part I: Theory," *Mech. Syst. Signal Process.*, vol. 24, no. 3, pp. 573–595, 2010.
[21] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.
[22] T. Söderström and P. Stoica. *System Identification*. Hemel Hempstead, Hertfordshire: Prentice Hall, 1989.
[23] P. C. Young, *Recursive Estimation and Time-Series Analysis. An Introduction for the Student and the Practitioner*. New York: Springer, 2011.
[24] K. Godfrey, "Design and application of multifrequency signals," *Comput. Control Eng. J.*, pp. 187–195, 1991.
[25] K. R. Godfrey, Ed., *Perturbation Signals for System Identification*. Upper Saddle River, NJ: Prentice Hall, 1993.
[26] K. R. Godfrey and M. Devenish, "An experimental investigation of continuous gas chromatography using pseudo-random-binary sequence," *Meas. Control*, vol. 2, no. 6, pp. 228–232, 1969.
[27] A. H. Tan and K. R. Godfrey, "The generation of binary and near-binary pseudorandom signals: An overview," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 4, pp. 583–588, 2002.
[28] K. R. Godfrey, A. H. Tan, H. A. Barker, and B. Chong, "A survey of readily accessible perturbation signals for system identification in the frequency domain," *Control Eng. Pract.*, vol. 13, no. 11, pp. 1391–1402, 2005.
[29] J. Schoukens, R. Pintelon, E. Vanderouderaa, and J. Renneboog, "Survey of excitation signals for FFT based signal analyzers," *IEEE Trans. Instrum. Meas.*, vol. 37, no. 3, pp. 342–352, 1988.
[30] J. Schoukens, M. Vaes, and R. Pintelon, "Linear system identification in a nonlinear setting: Nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation," *IEEE Control Syst. Mag.*, vol. 36, pp. 38–69, June 2016.
[31] P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods-Basic methods for consistent module estimates," *Automatica*, vol. 49, no. 10, pp. 2994–3006, 2013.
[32] M. Gevers and A. S. Bazanella. "Identification in dynamic networks: Identifiability and experiment design issues," in *Proc. 54th IEEE Conf. Decision and Control*, Osaka, Japan, Dec. 15–18, 2015, pp. 4005–4010.
[33] J. Schoukens, R. Pintelon, and H. Van Hamme, "Identification of linear dynamic systems using piecewise-constant excitations: Use, misuse and alternatives," *Automatica*, vol. 30, no. 7, pp. 1153–1169, 1994.
[34] E. O. Brigham, *The Fast Fourier Transform*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[35] K. J. Åström, *Introduction to Stochastic Control Theory*. New York: Academic, 1970.
[36] J. Schoukens, Y. Rolain, G. Simon, and R. Pintelon, "Fully automated spectral analysis of periodic signals," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1021–1024, 2003.
[37] W. Seifert, "Kommerzielle frequenzgangmesseinrichtungen," *Regelungstechnik*, vol. 10, no. 8, pp. 350–353, 1962.
[38] G. C. Anderson, B. W. Finnie, and G. T. Roberts, "Pseudo-random and random test signals," *Hewlett-Packard J.*, vol. 19, pp. 2–14, 1967.
[39] M. M. Levy, "Fourier transform analysis," *J. Br. Inst. Radio Eng.*, pp. 228–246, Mar.-May 1946.
[40] P. E. A. Cowley, "The application of an analog computer to the measurement of process dynamics," *Trans. ASME*, vol. 79, no. 4, pp. 823–832, 1957.
[41] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Historical notes on the fast Fourier Transform," *Proc. IEEE*, vol. 55, no. 2, pp. 1675–1677, 1967.
[42] G. D. Bergland, "A guided tour of the fast Fourier transform," *IEEE Spectr.*, pp. 41–52, July 1969.
[43] D. M. Morrison, B. W. Finnie, R. S. Patel, and K. H. Edwards, "Versatile display unit extends correlator capability," *Hewlett-Packard J.*, vol. 24, pp. 8–15, Nov. 1972.
[44] D. E. Torfs, R. Vuerinckx, J. Swevers, and J. Schoukens, "Comparison of two feedforward design methods aiming at accurate trajectory tracking of the end point of a flexible robot arm," *IEEE Trans. Autom. Control Syst. Technol.*, vol. 6, no. 1, pp. 2–14, 1998.
[45] P. Guillaume, J. Schoukens, R. Pintelon, and I. Kollar, "Crest-factor minimization using nonlinear Chebyshev-approximation methods," *IEEE Trans. Instrum. Meas.*, vol. 40, no. 6, pp. 982–989, 1991.
[46] M. R. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," *IEEE Trans. Inf. Theory*, vol. 16, no. 1, pp. 85–89, 1970.
[47] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume, "Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets," *Automatica*, vol. 33, no. 6, pp. 1073–1086, 1997.
[48] K. Barbe, J. Schoukens, and R. Pintelon, "Frequency-domain, errors-in-variables estimation of linear dynamic systems using data from overlapping subrecords," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1529–1536, 2008.
[49] P. Guillaume, R. Pintelon, and J. Schoukens, "Nonparametric frequency-response function estimators based on nonlinear averaging techniques," *IEEE Trans. Instrum. Meas.*, vol. 41, no. 6, pp. 739–746, 1992.
[50] R. Pintelon and J. Schoukens, *System Identification. A Frequency Domain Approach*. Piscataway, NJ: IEEE Press, 2001.
[51] P. Guillaume, "Frequency response measurements of multivariable systems using nonlinear averaging techniques," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 3, pp. 796–800, 1998.
[52] R. Pintelon, Y. Rolain, and W. Van Moer, "Probability density function for frequency response function measurements using periodic signals," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 1, pp. 61–68, 2003.
[53] P. M. T. Broersen, "A comparison of transfer-function estimators," *IEEE Trans. Instrum. Meas.*, vol. 44, no. 3, pp. 657–661, 1995.
[54] W. P. Heath, "The variance of nonparametric errors-in-variables estimates," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 1, pp. 228–236, 2005.
[55] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.
[56] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communication engineering," *Bell Syst. Tech. J.*, vol. 37, pp. 183–282, Jan. 1958.
[57] A. Stenman, F. Gustafsson, D. E. Rivera, L. Ljung, and T. McKelvey, "On adaptive smoothing of empirical transfer function estimates," *Control Eng. Pract.*, vol. 8, no. 11, pp. 1309–1315, 2000.
[58] P. E. Wellstead, "Using digital spectral techniques," Solartron Instruments, Tech. Rep. 008/83, 1984.

[59] P. E. Wellstead, "Theory and statistical accuracy of spectral analysis," Solartron Instruments, Tech. Rep. 009/83, 1984.
[60] J. W. Cooley and J. W. Tukey, "An algorithm for machine calculation of complex Fourier series," *Math. Comput*., vol. 19, pp. 297–301, Apr. 1965.
[61] F. J. Harris, "Use of windows for harmonic-analysis with discrete Fourier-transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 1978.
[62] J. Schoukens, Y. Rolain, and R. Pintelon, "Analysis of windowing/leakage effects in frequency response function measurements," *Automatica*, vol. 42, no. 1, pp. 27–38, 2006.
[63] J. Antoni and J. Schoukens, "A comprehensive study of the bias and variance of frequency-response-function measurements: Optimal window selection and overlapping strategies," *Automatica*, vol. 43, no. 10, pp. 1723–1736, 2007.
[64] J. Schoukens, G. Vandersteen, R. Pintelon, E. Zlatko, and Y. Rolain, "Bounding the polynomial approximation errors of frequency response functions," *IEEE Trans. Instrum. Meas*., vol. 62, no. 5, pp. 1346–1353, 2013.
[65] P. Verboven, E. Parloo, B. Cauberghe, and P. Guillaume, "Improved modal parameter estimation for lowly damped systems using non-parametric exponential windowing techniques," *Mech. Syst. Signal Process*., vol. 19, no. 4, pp. 675–699, 2005.
[66] J. L. Douce, "A note on frequency-response measurement," *Proc. Inst. Elec. Eng. Control Theory Appl*., vol. 133, no. 4, pp. 189–190, 1986.
[67] W. D. Widanage, J. L. Douce, and K. R. Godfrey, "Effects of overlapping and windowing on frequency response function estimates of systems with random inputs," *IEEE Trans. Instrum. Meas*., vol. 58, no. 1, pp. 214–220, 2009.
[68] P. Cawley, "The reduction of bias error in transfer-function estimates using FFT-based analyzers," *Trans. ASME, J. Vib. Acoust. Stress Reliab. Des*., vol. 106, no. 1, pp. 29–35, 1984.
[69] J. L. Douce and L. Balmer, "Statistics of frequency-response estimates," *Proc. Inst. Elec. Eng. Control Theory Appl*., vol. 137, no. 5, pp. 290–296, 1990.
[70] R. Pintelon, J. Schoukens, and G. Vandersteen, "Frequency domain system identification using arbitrary signals," *IEEE Trans. Autom. Control*, vol. 42, no. 12, pp. 1717–1720, 1997.
[71] T. McKelvey, "Frequency domain identification methods," *Circuits Syst. Signal Process*., vol. 21, pp. 39–55, 2002.
[72] J. L. Douce, "Improving frequency-response estimates by reduction of end effects," *Proc. Inst. Elec. Eng. Control Theory Appl*., vol. 153, no. 2, pp. 247–250, 2006.
[73] J. Schoukens, R. Pintelon, and Y. Rolain, *Mastering System Identification in 100 Exercises*. Hoboken, NJ: Wiley, 2012.
[74] J. Schoukens, G. Vandersteen, Y. Rolain, and R. Pintelon, "Frequency response function measurements using concatenated subrecords with arbitrary length," *IEEE Trans. Instrum. Meas*., vol. 61, no. 10, pp. 2682–2688, 2012.
[75] E. Geerardyn, "Development of user-friendly system identification techniques," Ph.D. dissertation, Vrije Universiteit Brussel, Brussels, Belgium, 2016.
[76] P. Hagg, J. Schoukens, M. Gevers, and H. Hjalmarsson, "The transient impulse response modeling method for non-parametric system identification," *Automatica*, vol. 68, pp. 314–328, June 2016.
[77] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe, "Estimation of nonparametric noise and FRF models for multivariable systems. Part II: Extensions, applications," *Mech. Syst. Signal Process*., vol. 24, no. 3, pp. 596–616, 2010.
[78] E. C. Levy, "Complex-curve fitting," *IEEE Trans. Autom. Control*, vol. 4, pp. 37–43, May 1959.
[79] T. McKelvey and G. Guerin. "Non-parametric frequency response estimation using a local rational model," in *Proc. 16th Int. Federation Automatic Control Symp. System Identification*, Brussels, Belgium, July 11–13, 2012, pp. 49–54.
[80] J. Lataire and T. Chen, "Transfer function and transient estimation by Gaussian process regression in the frequency domain," *Automatica*, vol. 68, pp. 217–229, Oct. 2016.
[81] P. Thummala and J. Schoukens, "Estimation of the FRF through the improved local bandwidth selection in the local polynomial method," *IEEE Trans. Instrum. Meas*., vol. 61, no. 10, pp. 2833–2843, 2012.
[82] D. Ugryumova, R. Pintelon, and G. Vandersteen, "Frequency response function estimation in the presence of missing output data," *IEEE Trans. Instrum. Meas*., vol. 64, no. 2, pp. 541–553, 2015.
[83] D. Ugryumova, R. Pintelon, and G. Vandersteen, "Frequency response matrix estimation from missing input-output data," *IEEE Trans. Instrum. Meas*., vol. 64, no. 11, pp. 3124–3136, 2015.
[84] B. Sanchez, J. Schoukens, R. Bragos, and G. Vandersteen, "Novel estimation of the electrical bioimpedance using the local polynomial method. Application to in vivo real-time myocardium tissue Impedance characterization during the cardiac cycle," *IEEE Trans. Biomed. Eng*., vol. 58, no. 12, pp. 3376–3385, 2011.
[85] P. M. J. Van den Hof and R. J. P. Schrama, "An indirect method for transfer-function estimation from closed-loop data," *Automatica*, vol. 29, no. 6, pp. 1523–1527, 1993.
[86] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215–1241, 1999.
[87] B. Ninness and H. Hjalmarsson, "On the frequency domain accuracy of closed-loop estimates," *Automatica*, vol. 41, no. 7, pp. 1109–1122, 2005.
[88] B. Wahlberg, "System-identification using Laguerre models," *IEEE Trans. Autom. Control*, vol. 36, no. 5, pp. 551–562, 1991.
[89] P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg, Eds., *Modelling and Identification with Rational Orthogonal Basis Functions*. London: Springer, 2005.
[90] W. G. Halvorsen and D. L. Brown, "Impulse technique for structural frequency-response testing," *Sound Vib*., vol. 11, no. 11, pp. 8–21, 1977.
[91] V. Wilkens and C. Koch, "Amplitude and phase calibration of hydrophones up to 70 MHz using broadband pulse excitation and an optical reference hydrophone," *J. Acoust. Soc. Amer*., vol. 115, no. 6, pp. 2892–2903, 2004.
[92] A. Link, A. Taeubner, W. Wabinski, T. Bruns, and C. Elster, "Calibration of accelerometers: Determination of amplitude and phase response upon shock excitation," *Meas. Sci. Technol*., vol. 17, no. 7, pp. 1888–1894, 2006.
[93] H. Fueser, S. Eichstaedt, K. Baaske, C. Elster, K. Kuhlmann, R. Judaschke, K. Pierz, and M. Bieler, "Optoelectronic time-domain characterization of a 100 GHz sampling oscilloscope," *Meas. Sci. Technol*., vol. 23, no. 2, Feb. 2012.
[94] M. Kobusch, T. Bruns, L. Klaus, and M. Mueller, "The 250 kN primary shock force calibration device at PTB," *Measurement*, vol. 46, no. 5, pp. 1757–1761, 2013.
[95] K. R. Godfrey, *Compartmental Models and Their Application*. New York: Academic, 1983.
[96] J. DiStefano, *Dynamic Systems Biology Modeling and Simulation*. London: Academic, 2013.
[97] T. M. Souders, D. R. Flach, C. Hagwood, and G. L. Yang, "The effects of timing jitter in sampling systems," *IEEE Trans. Instrum. Meas*., vol. 39, no. 1, pp. 80–85, 1990.
[98] H. Rake, "Step response and frequency-response methods," *Automatica*, vol. 16, no. 5, pp. 519–526, 1980.
[99] G. C. Anderson and M. A. Perry, "A calibrated real-time correlator/averager/probability analyzer," *Hewlett-Packard J*., vol. 19, pp. 9–20, Nov. 1969.
[100] B. Dwyer, "A feedback transfer function analyser," *Control*, vol. 9, pp. 676–678, Dec. 1965.
[101] K.R. Godfrey. "Introduction to binary signals used in system-identification," in *Proc. Int. Conf. Control*, Edinburgh, Scotland, Mar. 25–28, 1991, vol. 1, pp. 161–166.
[102] H. K. Wong, J. Schoukens, and K. R. Godfrey, "Structured non-linear noise behaviour and the use of median averaging in non-linear systems with m-sequence inputs," *IET Control Theory Appl*., vol. 7, no. 7, pp. 997–1004, 2013.
[103] A. De Angelis, J. Schoukens, K. R. Godfrey, and P. Carbone, "Practical synthesis of ternary sequences for system identification," *IEEE Trans. Instrum. Meas*., vol. 66, pp. 212–222, Feb. 2017.
[104] K. R. Godfrey, "The theory of the correlation method of dynamic analysis and its application to industrial processes and nuclear power plant," *Meas. Control*, vol. 2, no. 5, pp. T65–T72, 1969.
[105] K. R. Godfrey, "Dynamic analysis of an oil-refinery unit under normal operating conditions," *Proc. Inst. Elect. Eng*., vol. 116, no. 5, pp. 879–888, 1969.
[106] S. L. Marple, "Efficient least-squares FIR system-identification, *IEEE Trans. Acoust., Speech, Signal Process*., vol. 29, no. 1, pp. 62–73, 1981.
[107] B. Ninness and H. Hjalmarsson, "Variance error quantifications that are exact for finite-model order," *IEEE Trans. Autom. Control*, vol. 49, pp. 1275–1291, Aug. 2004.
[108] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*. London: Prentice-Hall, 1997.
[109] A. Marconato, M. Schoukens, and J. Schoukens. "Filter-based regularisation for impulse response modelling," *IET Control Theory Appl*., vol. 11, no. 2, pp. 194–204, 2017.