User choices for nonparametric preprocessing in system identification ⋆

J.Schoukens, G. Vandersteen, M. Gevers, R. Pintelon Y. Rolain *

* *Vrije Universiteit Brussel, dept. ELEC, Pleinlaan 2, 1050 Brussel, Belgium. E-mail: johan.schoukens@vub.ac.be*

Abstract Most research on system identification is focused on the identification of parametric models, for example a transfer function or a state space model where the information is condensed in a few parameters. In the daily practice, nonparametric methods, like frequency response function measurements, are intensively used. Recently, it was indicated that nonparametric identification methods could be used to robustify the parametric identification framework. A nonparametric preprocessing step can also be used to reduce or even eliminate the required user interaction, making system identification accessible for a much wider user group. For that reason, there is an increasing interest in nonparametric identification. In order to choose, compare, and to benchmark these nonparametric methods, it is very important to select the proper criteria. In this paper we identify and discuss the important choices that should be considered. It will be shown that these strongly depend on the intended use of the nonparametric model.

Keywords: Nonparametric identification, bias-variance trade-off, smoothing

1. INTRODUCTION

The mainstream of system identification is focused on the identification of parametric models, either in the time domain or in the frequency domain (Ljung, 1999; Pintelon and Schoukens, 2001; Söderström and Stoica, 1989). Many success stories are reported in the literature, for example the prediction error framework delivers dedicated models like ARX, ARMAX, OE, or Box-Jenkins depending upon the user selected noise model.

In practical applications, the use of non-parametric models is also very popular. The measurement of the impulse response using correlation methods (Godfrey, 1980), or the frequency response function (FRF) using spectral methods (Bendat and Piersol, 1980) are even today intensively used in commercial equipment. The most important drawback of these classical solutions is that they suffer from leakage errors, even if there is no disturbing noise present on the data. The development of the local polynomial method (Pintelon *et. al.*, 2010 a, b; Schoukens *et. al.*, 2009) was the start of a new class of nonparametric methods that allow the leakage to be reduced to levels where it does no longer harm the results. In a second step, alternative methods were proposed that reduce the noise sensitivity of the nonparametric methods by smoothing the results over neighboring frequencies (Gevers *et. al.*, 2011, Hägg *et. al.*, 2011).

Untill now, the interaction between parametric and nonparametric identification methods is rather low. But recently it was shown that the availability of good nonparametric plant and noise models allow us to improve the results obtained in the parametric framework (Schoukens *et. al.*, 2011). Nonparametric results from a preprocessing step are not only used to guide the user during the model selection process, the availability of a nonparametric noise model can also be used to improve the robustness and the convergence of the parametric identification methods.

A series of new nonparametric methods are nowadays under study, and this creates a need to compare the older and more recent methods. What are the best nonparametric methods? There is no single answer to this question, the intended use of the model will strongly affect the discussion. In order to avoid an unstructured discussion with a proliferation of many solutions and answers, we need a framework to bring structure into this field. The aim of this paper is to address this need. We will identify a number of critical issues that will help to guide the researcher and the user to make a clear classification that will allow the methods to be compared to each other.

In Section 2 we introduce first the improved nonparametric FRF estimator for linear dynamic systems. The bias/variance tradeoff of smoothed estimates is discussed in Section 3. In Section 4 we reflect on the dependency of the optimal properties on the intended application. Next we illustrate the results on an example in Section 5 before we draw the final conclusions.

2. A BRIEF INTRODUCTION TO THE LPM, A NONPARAMETRIC FRF ESTIMATOR

For a long time, the estimation of the FRF started from the estimated cross- and auto-spectrum, or cross- and auto-correlation (Bendat and Piersol, 1980). The inherent leakage problems were reduced and reshaped by applying

⋆ This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), and by the Belgian Government through the Interuniversity Poles of Attraction (IAP VI/4) Program.

windowing techniques. Recently an alternative method, the local polynomial method (LPM), was proposed that estimates the frequency response function (FRF) $\hat{G}_{poly}$ and the power spectrum of the disturbing noise with a much higher quality than these classical windowing methods. The cost for this improved quality is an increased computation time (typical a factor 1000 independently of the record length). But, given the actual available computer power, records of a few (ten)thousands of data points are still processed in a few seconds. The required calculation time grows proportionally with the record length.

Consider the system given by

$$y\left(t\right)=G_0\left(q\right)u_0\left(t\right)+v\left(t\right), \tag{1}$$

where q^{-1} is the backward shift operator, and $v\left(t\right)$ the disturbing noise modeled as filtered white noise: $v\left(t\right)=H_0\left(q\right)e\left(t\right)$. For a finite record length

$$u_0\left(t\right), y\left(t\right)\ \ t=0,\ldots,N-1$$

this equation has to be extended with the initial conditions (transient) effects of the dynamic plant and noise system t_G, t_H:

$$y\left(t\right)=G_0\left(q\right)u_0\left(t\right)+H_0\left(q\right)e\left(t\right)+t_G\left(t\right)+t_H\left(t\right), \tag{2}$$

with $t=0,\ldots,N-1$. Using the discrete Fourier transform (DFT)

$$X(k)=\frac{1}{\sqrt{N}}\sum_{t=0}^{N-1}x(t)e^{-j2\pi kt/N} \tag{3}$$

an exact frequency domain formulation of (2) is obtained:

$$\begin{aligned}Y\left(k\right)&=G_0\left(\Omega_k\right)U_0\left(k\right)+T_G\left(\Omega_k\right)+H_0\left(\Omega_k\right)E(k)+T_H\left(\Omega_k\right)\\&=G_0\left(\Omega_k\right)U_0\left(k\right)+H_0\left(\Omega_k\right)E(k)+T\left(\Omega_k\right)\end{aligned} \tag{4}$$

where the index k points to the frequency kf_s/N with f_s the sampling frequency. The contributions U_0, E, Y in (4) are an $O\left(N^0\right)$, the transient terms T_G, T_H are an $O\left(N^{-1/2}\right)$, where $O(x)$ = ordo x: a function tending to zero at the same rate as x. It is most important for the rest of this paper to understand that (4) is an exact relation (Pintelon *et. al.*, 2010 a,b; Pintelon and Schoukens, 2001; McKelvey, 2000). The finite length record requires the use of a transient term in (2), and it turns out that the leakage errors of the DFT are modeled by very similar terms in the frequency domain. All these terms $t_G\left(t\right), t_H\left(t\right), T_G\left(\Omega_k\right), T_H\left(\Omega_k\right)$ are described by rational forms in q^{-1} (time domain) or z^{-1}(frequency domain). The presence of the transient or leakage terms is independent of the selected domain to describe the system (time- or frequency domain), but it depends on the nature of the excitation (periodic or random). A finite length experiment will always suffer from 'leakage' errors as well in time- as in frequency domain.

Making use of the smoothness of G_0, H_0 and T, the following polynomial representation holds for the frequency lines $k+r$, with $r=0,\pm1,\ldots,\pm n$.

$$\begin{aligned}&G_0(\Omega_{k+r})=\\&G_0(\Omega_k)+\sum_{s=1}^{R}g_k(s)r^s+O(\Big(\frac{r}{N}\Big)^{(R+1)})\end{aligned} \tag{5}$$

$$\begin{aligned}&T_G(\Omega_{k+r})=\\&T(\Omega_k)+\sum_{s=1}^{R}t_k(s)r^s+N^{-\frac{1}{2}}O(\Big(\frac{r}{N}\Big)^{(R+1)})\cdot\end{aligned} \tag{6}$$

Putting all parameters $G_0(\Omega_k), T_G(\Omega_k)$ and the parameters of the polynomial $g_k\left(s\right), t_k(s)$, with $s=1,\ldots,R$ in a column vector θ_k, and their respective coefficients in a row vector, $K(k,r)$ allows (4) to be rewritten (neglecting the remainders) as:

$$Y(k+r)=K(k,r)\theta_k, \tag{7}$$

Collecting (7) for $r=-n,-n+1,\ldots,0,\ldots,n$ finally gives

$$Y_{n,k}=K_{n,k}\theta_k,\text{ for }k=0,\ldots,N/2-1 \tag{8}$$

with $Y_{n,k}, K_{n,k}$ the values of $Y(k+r), K(k,r)$, for $r=-n,-n+1,\ldots,0,\ldots,n$ stacked on top of each other. Observe that the matrix $K_{n,k}$ depends upon U_0. Solving this equation in a least squares sense eventually provides the polynomial least squares estimate $\hat{G}_{\text{poly}}(\Omega_k)$ for $G(\Omega_k)$. From the residuals of the fit, an estimate $\hat{\sigma}_V^2\left(k\right)$ of $\sigma_V^2\left(k\right)=\left|H(\Omega_k)\right|^2\sigma_e^2\left(k\right)$ is obtained. In order to get a full rank matrix K_n, enough spectral lines should be combined: $n\geq R+1$. The smallest interpolation error is obtained for $n=R+1$.

A detailed discussion of these results can be found in (Pintelon *et. al.*, 2010 a,b; Schoukens *et. al.*, 2009).).

3. THE BIAS/VARIANCE TRADE-OFF OF FRF ESTIMATORS

Just as for any identification method, we have to make a bias-variance trade-off when evaluating the quality of the nonparametric FRF estimates. In this case we will consider the situation where the variance induced by the disturbing noise $v\left(t\right)$ dominates the variance that is due to the leakage error (see Pintelon *et. al.* 2010 (a,b) for more details). It is possible to modify the bias/variance ratios either by changing the settings of the LPM, or by adding an additional smoothing to the algorithm. Both methods are briefly discussed.

3.1 Impact of the settings of the LPM on the bias/variance trade off

The LPM method is controlled by two parameters. The first is the degree R of the local polynomials, the second is the local bandwidth that is specified by n. The bias and variance on the nonparametric estimate $\hat{G}_{\text{poly}}(\Omega_k)$is given by respectively by (Pintelon *et al.*, 2010a):

$$G_B(\Omega_k)=O((n/N)^{R+1} \tag{9}$$

and

$$\sigma_G^2\left(k\right)=O(\sigma_V^2(k)/\left(2n-R\right))=O(\sigma_V^2(k)/n). \tag{10}$$

For the variance expression, we assumed that the disturbing noise is dominating the leakage error. This shows that the bias/variance ratio is affected by the choice of n, R. Restricting n to its minimum value ($n=R+1$) will result in the smallest bias, but a higher variance. Increasing n will reduce the variance at a cost of a fast increasing bias. When the variance is balanced against the bias ($\sigma_G^2\left(k\right)=\left|G_B(\Omega_k)\right|^2$) we have that

$$\tilde{n} = O(\left(\sigma_V^2\right)^{\frac{1}{2R+3}} N^{\frac{2R+2}{2R+3}}) \tag{11}$$

and

$$\tilde{G_B^2} = \tilde{\sigma}_G^2 = O(\left(\sigma_V^2\right)^{\frac{2R+2}{2R+3}} N^{-\frac{2R+2}{2R+3}}) \tag{12}$$

For $R = 2$ we find that $\tilde{G_B^2} = \tilde{\sigma}_G^2 = O(\left(\sigma^2\right)^{\frac{6}{7}} / N^{\frac{6}{7}})$, and $\tilde{n} = O(\sigma^{\frac{2}{7}} N^{\frac{6}{7}})$.

These results show clearly that the user can tune the LPM either towards a minimal bias (at a cost of an increased variance), or a balanced bias/variance solution. In the next section we introduce smoothing as an alternative for changing the parameters of the LPM.

3.2 Imposing additional smoothing

The basic nonparametric methods process the data frequency by frequency. The results over the different frequencies are not linked to each other although we know that the FRF is a smooth function of the frequency. Imposing additional smoothness constraints to the solution by combining the neighbouring frequencies results in a noise reduction. Smoothing methods with a variable smoothing window are studied in Stenman and Gustafsson (2001) and Fan and Gijbels (1995). Recently, smoothing was also added to the local polynomial method (Gevers *et. al.*, 2011). Imposing smoothing constraints over the neighbouring frequencies can be interpreted as if the local models that are used in the nonparametric method should describe a wider local frequency bandwidth. Hence we can use again the relations (12), but have to replace n by a new value $n_{eff} \geq n$ where n_{eff} will depend upon the nature and the settings of the smoothing method. An alternative method that estimates the FRF $\hat{G}(\Omega_k)$ in the frequency domain and the leakage terms $t_G(t)$in the time domain through a global Least Squares problem was recently introduced in (Hägg *et. al.*,2011) and compared with other nonparametric methods in Gevers *et. al.* (2012). It also introduces smoothing on the estimates $\hat{G}(\Omega_k)$ in order to produce enough equations for the LS problem.

3.3 Conclusion

We can balance the bias/variance trade-off either by adding smoothing constraints or by changing the settings of the LPM. For both choices it is clear that the variance is reduced at a cost of an increased bias in the results: increasing the smoothing by combining measurements over a wider frequency band increases also the systematic errors because the simple nonparametric or local models should span a wider frequency band.

4. OPTIMAL TUNING OF THE NONPARAMETRIC IDENTIFICATION METHODS

From the previous section, it turns out that the user can tune the bias-variance ratio of the nonparametric identification methods. In this section we will explain that the best ratio will depend upon the intended use of the nonparametric results. The estimation of the FRF of a system is only an intermediate step in a complex modelling process. In some applications the results are used to make a graphical representation of the dynamics of a system, while in other problems the FRF measurements (or even better, the transient compensated output Fourier transform) are used as a starting point for a parametric modeling step.

4.1 Direct use of the nonparametric results

If the nonparametric FRF measurements are directly interpreted, the user prefers a smooth result that is disturbed as less as possible by the noise. From such a figure it is for example possible to get an idea of the dominant frequency range of the system; the damping of the dominant system poles in the frequency range of interest; the phase (margin) of the system; etc. We can translate this desire in a formal requirement by minimizing the mean square error: the sum of the squared bias and variance should be as small as possible. This requires the variance to be balanced against the bias errors.

Besides an estimate of the plant model, the unsmoothed nonparametric method delivers also an estimate of the power spectrum of the disturbing noise. In the classical smoothing methods all emphasis is put on the bias/variance trade-off of the plant model, without paying any attention to the quality of the nonparametric estimate of the noise model. The noise model is estimated from the residuals of the fit. It is clear that it will be biased because also the bias errors of the plant model will contribute to the residuals. The power of this contribution at frequency Ω_k is given by:

$$G_B\left(\Omega_k\right)^2 S_U\left(k\right) \tag{13}$$

with $S_U\left(k\right)$ the power spectrum of the input at frequency k. When the nonparametric use of the plant model is the main goal, this is not an issue. However this is unacceptable when we need reliable uncertainty bounds, or reliable nonparametric noise model estimates. A two step method should be used in that case to provide the uncertainty bounds. In the first step, the smoothed plant model is estimated, in the second step (smoothing switched off) the noise model is obtained and next it is translated in uncertainty bounds (and if possible also an estimate of the bias) on the smoothed results.

4.2 Parametric identification starting from intermediate nonparametric results

A parametric model combines the measurements at all frequencies in one single model that uses significantly less parameters than the nonparametric FRF description. Also this step smoothes the nonparametric data. However, if a bias is present in the raw data used for the parametric identification, it will be impossible to remove it. Because many more data points per parameter are combined in the parametric identification step the variance that is induced by the noise will be much smaller, and the optimal balance between variance and bias errors should shift towards a significantly smaller bias. This is the opposite of what would be obtained by adding additional smoothing to the nonparametric estimates. For that reason no smoothing should be added to the nonparametric estimates if these will be used as an input for a parametric estimation step.

4.3 Conclusion

From the previous discusions it turns out that the intended goal of the nonparametric processing steps will set the required properties of the algorithm. If a high quality nonparametric plant model is the major goal, it is advisable to add smoothing to the algorithm. Alternatively, if the results will be used as the input of a parametric estimation step, or if the estimation of the nonparametric noise model is the goal of the modeling process, it is not advisable to add smoothing to the nonparametric preprocessing step. In that case most attention should be paid to the reduction of the bias of the nonparametric estimates.

5. EXAMPLE

The results of the previous discussions are illustrated on a simulation. Consider the following system:

$$G_0(z) = B_0(z)/A_0(z)$$

with

$$a = [0.19427 \quad 0.38854 \quad 0.19427]$$

and

$$b = [1 \quad 0.71246 \quad 0.74486].$$

The system is disturbed by process noise that is described by the following noise model

$$H_0(z) = C_0(z)/D_0(z)$$

with

$$c = [0.038854 \quad 0.083670 \quad 0.10156 \quad 0.059723 \quad 0.018658].$$

and

$$d = [1 \quad 1.5281 \quad 2.2864 \quad 1.2918 \quad 0.71537].$$

The system is excited with filtered white noise

$$u_0(t) = G_{gen}(q)e_u(t)$$

with

$$G_{gen}(q) = B_{gen}(q)/A_{gen}(q)$$

and

$$b_{gen} = [0.52762 \quad 1.5829 \quad 1.5829 \quad 0.52762]$$

and

$$a_{gen} = [1 \quad 1.7600 \quad 1.1829 \quad 0.27806].$$

The simulation is repeated 10 000 times. In each simulation $N + 500$ data points were generated. The first 500 points of each simulation were removed in order to create simulations with non-zero initial conditions. The remaining $N = 2000$ data points were processed to estimate the (smoothed) FRF and the variance of the noise. The rms error on the FRF (see Fig. 1), and the mean of the estimated variance (see Fig. 3) is calculated over the 10 000 realizations. The number of realizations is selected so high in order to show very clearly the bias/variance trade off. From the averaged FRF estimates, the bias on the estimated FRF is calculated (see Fig.2). From the figures, it is seen that the rms error drops in this simulation when the smoothing is increased (the local bandwidth is increased). This is to be expected since the disturbing noise levels were much higher than the leakage errors of the LPM. In Fig. 2 it can be seen that at the same time also the bias increased. For the largest and 2nd largest local bandwidth a clear bias can be seen. For the smallest bandwidth the bias is

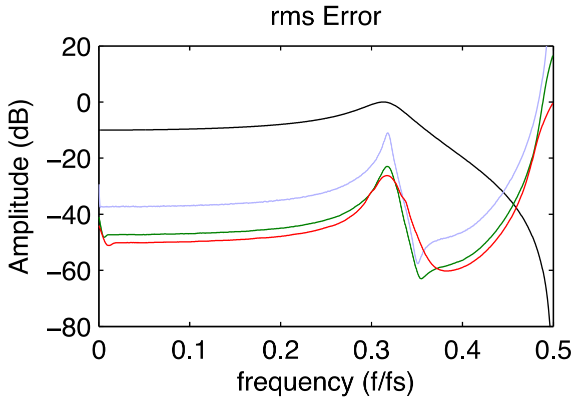

Figure 1. rms error of the estimated FRF for different local bandwidths of the LPM. Black: G_0; blue: $n = 3$; green: $n = 19$; red $n = 38$.

still below the noise level, even after averaging over 10 000 realizations. This is the reason that the blue curve is not smooth, the noise error still dominates the results. In the last figure it can be seen that the variance estimate that is obtained with the smallest local bandwidth coincides completely with the true value, while for the larger local bandwidths a bias can be observed: the sharp resonance is smeared over the neighbouring frequencies, and the anti-resonance is lost.

These results are in perfect agreement with the earlier discussions.

i) If the nonparametric estimate of the FRF will be used the largest local bandwidth should be selected. It results in the smallest rms-error and hence the best results to be used for a direct interpretation.

ii) If the estimated noise variance will be further used as a frequency weighting in a parametric estimate, we have to avoid the bias that is apparent for the smoothed results. This would lead to a loss in efficiency on the parametric estimate. Also the generation of uncertainty bounds would be badly influenced by these wrong estimates.

iii) If in addition also the nonparametric FRF estimate (or the transient compensated input/output Fourier transform) would be used in a second step to obtain a parametric plant model estimate the situation becomes even worse. In that case also a bias will be created on the parametric estimate.

Conclusion: these results show clearly that the selection of the best nonparametric method depends on the intended use of the results. As a rule of thumb, we can say that unsmoothed results should be used if the nonparametric estimates are used as the input for a parametric post processing. If the nonparametric FRF is the final goal (to determine for example the cross-over frequency) it is better to add an additional smoothing in order to get the smallest rms-errors.

6. CONCLUSION

In this paper we set up a framework to compare and select the 'best' nonparametric FRF-estimator. It is shown that the optimal choice depends strongly on the intended use of

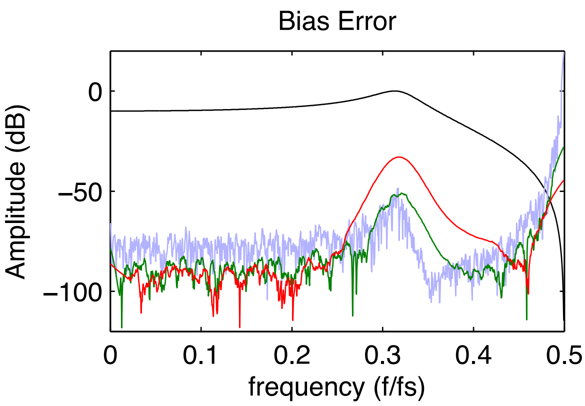

Figure 2. Bias error on the estimated FRF for different local bandwidths of the LPM. Black: G_0; blue: $n = 3$; green: $n = 19$; red $n = 38$.

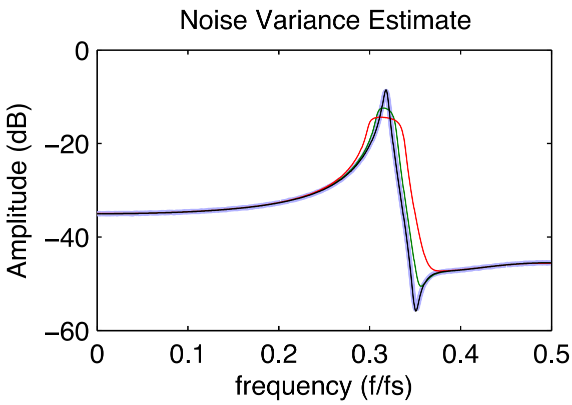

Figure 3. Estimated variance of the disturbing noise for different local bandwidths of the LPM. Black: true variance; Estimated variance: blue: $n = 3$; green: $n = 19$; red $n = 38$.

the results, and upon the experimental conditions. If the nonparametric results are directly used and interpreted by the user an optimized smoothing should be included in the nonparametric estimate. The smoothing reduces the variance of the estimate, but creates at the same time a bias that increases with an increasing smoothing. The bias/variance trade off can be tuned in order to minimize the rms-error of the FRF-estimate. The smaller the disturbing noise (with respect to the level of the leakage errors), the less smoothing should be applied. If the nonparametric estimates (noise model or/and plant model) are to be used later in a parametric estimate, it is important to avoid a bias because this can not be removed any more in the post-processing. For that reason smoothing should be turned off in that situation.

ACKNOWLEDGEMENTS

Vrije Universiteit Brussel, Department ELEC. This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), and by the Belgian Government through the Interuniversity Poles of Attraction (IAP VI/4) Program. ?

7. REFERENCES

Bendat J.S., . and A.G. Piersol (1980). *Engineering Applications of Correlations and Spectral Analysis.* Wiley, New York.

Fan J.Q. and Gijbels I (1995). “Data-driven bandwidth selection in local polynomial fitting - variable bandwidth and spatial adaptation,” *Journal of the Royal Statistical Society Series B-Methodological*; 57, pp. 371-394.

Gevers M., R. Pintelon, and J. Schoukens (2011). The local polynomial method for nonparametric system identification: improvements and experimentation. Accepted for presentation at the *50th IEEE Conference on Decision Control and 10th European Control Conference*, Orlando-Florida(USA), Dec. 12-15.

Gevers M., P. Hägg, H. Hjalmarsson, R. Pintelon, and J. Schoukens (2012). Comparison between the KTH method and the Local Polynomial Method for nonparametric system identification. Submitted to the *16th IFAC Symposium on System Identification*, Brussels (Belgium), July 11-13.

Godfrey K.R. (1980) Correlation Methods. *Automatica*, 16, 527-534.

Hägg P., H. Hjalmarsson, and B. Wahlberg (2011). A least squares approach to direct frequency response estimation. Accepted for presentation at the *50th IEEE Conference on Decision Control and 10th European Control Conference*, Orlando-Florida(USA), Dec. 12-15.

Ljung L. (1999). System Identification: *Theory for the User (second edition).* Prentice Hall, Upper Saddle River, New Jersey.

McKelvey T. (2000). On the finite length DFT of input-output signals of multivariable linear systems. *In 39th IEEE Conference on Decision and Control.* Sidney (Australia), Dec. 12-15, pp. 5190-5191.

Pintelon R., and J. Schoukens (2001), *System Identification: A Frequency Domain Approach.* Piscataway (USA): IEEE Press.

Pintelon R., J. Schoukens, G. Vandersteen, and K. Barbé (2010a). Estimation of nonparametric noise and FRF models for multivariable systems - part I: theory, *Mechanical Systems and Signal Processing*, 2010; 24, pp. 573-595.

Pintelon R., J. Schoukens, G. Vandersteen, and K. Barbé (2010b). Estimation of nonparametric noise and FRF models for multivariable systems - part II: extensions, applications, *Mechanical Systems and Signal Processing*, 2010; 24, pp. 596-616.

Pintelon R., J. Schoukens, and P. Guillaume (2006). Continuous-time noise modelling from sampled data, *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 6, pp. 2253-2258.

Schoukens J., Vandersteen G., Barbe K., and R. Pintelon (2009). “Nonparametric Preprocessing in System Identification: a Powerful Tool,” *European Journal of Control*; 15, pp. 260-274.

Schoukens J., Y. Rolain, G. Vandersteen, and R. Pintelon (2011). User friendly Box-Jenkins identification using nonparametric noise models. Accepted for presentation at the *50th IEEE Conference on Decision Control and 10th European Control Conference*, Orlando-Florida(USA), Dec. 12-15.

Söderström T. and Stoica P. (1989). *System Identification.* Prentice-Hall, Englewood Cliffs.

Stenman A., Gustafsson F. (2001). Adaptive smoothing methods for frequency-function estimation, *Automatica*; 37, pp. 675-685.