2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

User friendly Box-Jenkins identification using nonparametric noise models

J. Schoukens, Y. Rolain, G. Vandersteen, R. Pintelon

***Abstract*—The identification of SISO linear dynamic systems in the presence of output noise disturbances is considered. A 'nonparametric' Box-Jenkins approach is studied: the parametric noise model is replaced by a nonparametric model that is obtained in a preprocessing step, and this without any user interaction. The major advantage for the user is that i) one method can be used to replace the classical ARX, ARMAX, OE, and Box-Jenkins models; ii) no noise model order should be selected. This makes the identification much easier to use for a wider public; iii) a bias on the plant model does not create a bias on the noise model. The disadvantage of the proposed nonparametric approach is a small loss in efficiency with respect to the optimal parametric choice. These results are illustrated on a series of well selected problems.**

***Index Terms*—system identification, non-parametric noise models, Box-Jenkins**

I. Introduction

In the classical time domain prediction error framework, a parametric plant- and noise model is estimated simultaneously for the system given by

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

where q^{-1} is the backward shift operator, and with $v(t)$ the disturbing noise modeled as filtered white noise: $v(t) = H_0(q)\, e(t)$. The plant and noise models are respectively given by

$$G_0(q) = B_0(q) / A_0(q),$$

and

$$H_0(q) = C_0(q) / D_0(q),$$

with A_0, B_0, C_0, D_0 polynomials in q. During the identification step, the noise model $H(q, \theta)$ acts as a parameter dependent filter on the residuals in the least squares cost function [2], [3]

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \left(H^{-1}(q, \theta) \left[y(t) - G(q, \theta)\, u_0(t) \right] \right)^2, \tag{2}$$

which adds a frequency weighting to the cost function. The user has to select the model structure of both the plant model $G(q, \theta)$ and the noise model $H(q, \theta)$. The choice of the noise model not only reflects the prior knowledge about the system, it also affects the complexity of the optimization problem to find the minimum of the cost function $V_N(\theta)$. For example, choosing $H(q, \theta) = 1/A(q, \theta)$ expresses that the plant and the noise models have the same poles, resulting in an optimization problem that is linear-in-the-parameters. This is called the ARX model. In the ARMAX model, there is a larger flexibility in the noise model by adding also zeros to the noise model: $H(q, \theta) = C(q, \theta) / A(q, \theta)$, Now, a nonlinear optimization problem is faced to minimize the cost function. For the output error model (OE), the disturbing noise is assumed to be white: $H(q, \theta) = 1$, and in the Box-Jenkins model there is no relation between the plant and the noise model [2], [3]. It is clear that the Box-Jenkins model can cover the ARX, ARMAX, and OE situation, but it results also in a more difficult optimization problem to be solved. The user has to solve now a double model selection problem: the order of both the plant- and the noise the model should be selected.

The noise model structure selection problem can be avoided if a good nonparametric noise model would be available. It can be used as a parameter independent weighting vector in the weighted least squares method. So only the plant model order has to be retrieved by the user. The numerical search procedure becomes also more robust so that the risk to end in local minima is reduced.

This brings us to the contribution of this paper. When dealing with system identification we can consider on the one hand the classical prediction error framework that makes use of parametric noise models. It results in optimal estimates (consistent and efficient), provided that the user makes the correct choices for the plant- and noise model structure and order. However, if the user fails to do so, these highly desirable properties are lost and (large) errors can be created. On the other hand we have the nonparametric noise model approach, where no user interaction at all is requested to select the noise model-structure and -order. Only the plant model structure selection should be addressed. This results in a very user friendly modeling technique at a cost of a loss in efficiency. However, the risk to end up with poor models due to a bad user choice is strongly reduced. So there is a possibility to trade optimal, but high risk methods, for good (not optimal), but low risk methods. In this paper we will

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.

study both approaches and discuss the needed trade-off.

In Section II we first extend the prediction error framework with a nonparametric noise model. Next, it is discussed how a nonparametric noise model can be obtained without user interaction from the data alone. Section III, studies the properties of the nonparametric Box-Jenkins model. The similarities and differences with earlier published results on nonparametric Box-Jenkins identification is discussed in Section IV. Next, a detailed comparison of ARX, OE, Box-Jenkins and nonparametric Box-Jenkins identification is made in Section V followed by the conclusions in Section VI.

II. Introducing a nonparametric noise model to the prediction error framework

In this section we show very briefly how a nonparametric noise model can be introduced in the classical prediction error framework. We refer the interested reader also to [7], [8], [9], [10].

A. Including a nonparametric noise model in the prediction error framework

For finite data lengths, the model in (2) should be extended with a transient term $t(\theta)$ to model the effects of the initial conditions of the plant and noise model:

$$\tilde{V}_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(H^{-1}(q,\theta)\left[y(t) - G(q,\theta)u_0(t) - t(\theta)\right]\right)^2. \quad (3)$$

The cost function (3) remains the same in the frequency domain using Parseval's theorem:

$$\tilde{V}_N(\theta) = \frac{1}{N}\sum_{k=-N/2+1}^{N/2} \left(H^{-1}(k,\theta)\left[Y(k) - G(k,\theta)U_0(k) - T(k,\theta)\right]\right)^2 \quad (4)$$

where e.g. $G(k,\theta)$ is the parametric plant model evaluated at the frequency $\Omega_k = e^{j2\pi k/N}$. The expressions (3) and (4) are completely equivalent (excepted for the transients of the noise filter $H^{-1}(q,\theta)$ that are neglected here for simplicity), so that it is even not a problem to use the same symbol for both expressions. Under these conditions, the time domain or frequency domain implementation are only two different ways to calculate the same result. The estimated plant model will be exactly the same. The estimated term $t(\theta)$ and $T(\theta)$ differ from each other because they have a slightly different role. In the time domain, $t(\theta)$ models the initial transients. In the frequency domain, $T(\theta)$ is an exact model for the leakage errors that can be written in the time domain as a transient at the beginning and the end of the measurement [4].

If a good nonparametric noise model

$$\hat{\sigma}_v^2(k) \approx \sigma_e^2 \left|H_0(k,\theta_0)\right|^2 \quad (5)$$

is available, the cost function (4) can be replaced by

$$\tilde{V}_{N,nonpar}(\theta) = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} \frac{\left|Y(k) - G(k,\theta)U_0(k) - T(k,\theta)\right|^2}{\hat{\sigma}_v^2(k)}, \quad (6)$$

where only the plant model $G(k,\theta)$ and the transient term $T(k,\theta)$ remain to be estimated. There is no interference any more with the estimation of the plant model so that a single model selection problem remains.

B. Estimation of a nonpararmetric noise model with the local polynomial method

We will discuss two methods to estimate the nonparametric noise model, i) the classical windowing method, and ii) the recently developed 'local polynomial method'.

Estimating the noise power spectrum using the windowing method: The most popular method to estimate the frequency response function $G_0(k)$ and the variance $\sigma_v^2(k)$ is based on the use of the estimated cross- and auto-spectrum of the input and output [11], [13]: $\hat{S}_{YU}(k)$, $\hat{S}_U(k)$, $\hat{S}_Y(k)$, usually estimated by making use of a Hanning window in order to reduce the leakage errors. The estimated variance with the Hanning window, $\hat{\sigma}_{v,H}^2(k)$, is given by:

$$\hat{\sigma}_{v,H}^2(k) = \hat{S}_Y(k) - \left|\hat{S}_{YU}(k)\right|^2 / \hat{S}_U(k). \quad (7)$$

The disadvantages of this approach are that the original record is split in M sub-records to calculate the cross- and auto-spectrum estimates which leads to a loss in frequency resolution [11] and increased leakage errors (bias contributions). It can be shown that [8], [9] for the Hanning window:

$$\hat{\sigma}_{v,H}^2(k) = \sigma_v^2(k) + \varepsilon_{H,1} + \varepsilon_{H,2} + \varepsilon_{H,3}, \quad (8)$$

with

i) $\varepsilon_{H,1} = \sigma_v^{2'}(k)\, O\left((M/N)^2\right)$: the interpolation error ($\sigma_v^{2'}(k)$ is the derivative of the power spectrum with respect to the frequency evaluated at k). The Hanning-method assumes that the noise variance is constant over 3 neighboring frequencies, resulting in an interpolation error that is localized in frequency [7].

ii) $\varepsilon_{H,2} = G_0^{'}(k)\, O\left((M/N)^2\right)$: the error due to a bad separation of $y_0(t)$ and $v(t)$. This is due to estimation errors on $G_{Hanning}(k)$, and it is usually the dominating error, especially for systems with fast varying dynamics.

iii) $\varepsilon_{H,3} = O\left((M/N)^3\right)$ is the error due to the remaining leakage after applying the Hanning window. As shown in Figure 1, it does carry over errors from one frequency band (for example with a high noise level) to another band (for example with a low noise level). That effect makes the leakage errors a much more disturbing problem in noise analysis than the interpolation error. For example, the interpolation error will be very small at zeros of the noise model, such that the leakage error becomes dominant over the interpolation error.

For slowly varying spectra, the interpolation error will be small because $\sigma_v^{2'}(k)$ goes to zero [7].

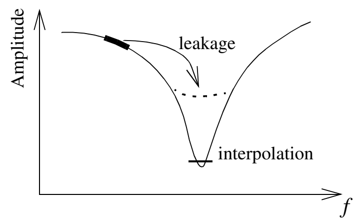

Figure 1. Illustration of the different nature of the leakage error and the interpolation error another).

Estimating the noise power spectrum using the local polynomial method: Recently an alternative method, $\hat{G}_{poly}$ (the local polynomial method) is proposed that estimates the frequency response function (FRF) and the power spectrum of the noise with a much higher quality than the classical windowing methods [11], at a cost of an increased computer time (typical a factor 1000 independently of the record length). But, given the actual available computer power, records of a few thousands of data points are still processed in a few seconds. The required calculation time grows proportionally with the record length. The method starts from the observation that

$$y(t) = G_0(q)\,u_0(t) + t_G(t) + H_0(q)\,e(t) + t_H(q)\,. \quad (9)$$

or in the frequency domain (using the discrete Fourier transform)

$$\begin{aligned} Y(k) &= G_0(k)\,U_0(k) + T_G(k) + H_0(qk) + T_H(k) \\ &= G_0(k)\,U_0(k) + H_0(qk) + T(k)\,. \end{aligned} \quad (10)$$

The plant-, noise-, and transient models are smooth functions of the frequency. Around a given frequency, they can be locally approximated by a complex polynomial that is estimated by solving a linear least squares problem at each frequency [7]. This leads eventually to a high quality estimate of the FRF $\hat{G}_{poly}$, the 'leakage corrected' output

$$\hat{Y}(k) = Y(k) - \hat{T}(k)\,. \quad (11)$$

The variance of the noise is estimated from the residuals of the fit [7], [8]:

$$\hat{\sigma}^2_{v,poly}(k) = \sigma_v^2(k) + \varepsilon_{H,1} + \varepsilon_{H,2} + \varepsilon_{H,3}. \quad (12)$$

with

i) $\varepsilon_{H,1} = \sigma_v^{2'}(k)\,O((1/N))$: the interpolation error. It should be remarked that with the local polynomial method a split of the data in sub-records is not required, so $M = 1$ which reduces all the error levels significantly. Compared to the Hanning-method there is a loss of one order of magnitude with respect to the decrease in N. However, as explained in Figure 1, this error is not the dominating error term in most practical situations [7].

ii) $\varepsilon_{H,2} = G_0^{(3)}(k)\,O\left((1/N)^6\right)$: the error due to a bad separation of $y_0(t)$ and $v(t)$. This was the dominating error term in most situations for the Hanning method. Observe that this term is strongly reduced by using the local polynomial method, and it is one of the reasons that the latter becomes superior compared to the Hanning approach in practice.

iii) $\varepsilon_{H,3} = O\left((1/N)^4\right)$ is the error due to the remaining leakage after applying the local polynomial method.

A detailed discussion of these results can be found in [8], [9], [7].

Using the nonparametric noise model in the prediction error framework: In this paper, we are mainly interested in the leakage corrected output estimate $\hat{Y}(k)$ (11) and its variance estimate $\hat{\sigma}^2_Y(k)$. This variance is estimated completely similar to $\hat{\sigma}^2_{v,poly}(k)$ using the local polynomial method. This allows us to reformulate (13) in its final form:

$$\tilde{V}_{N,nonpar}(\theta) = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} \frac{\left|\hat{Y}(k) - G(k,\theta)U_0(k)\right|^2}{\hat{\sigma}^2_Y(k)}. \quad (13)$$

So we are now in the situation that we have a free choice between using the parametric- (4) or the nonparametric noise model approach (13). In the latter expression $U(k), \hat{Y}(k)$, (leakage + initial transient elimination) and $\hat{\sigma}^2_G(k)$ (nonparametric noise model) are used during the estimation. It is on these choices that we will further elaborate in this paper. In the rest of this paper we will call the nonparametric noise model formulation (13) the nonparametric Box-Jenkins approach because of the absence of a link between the plant- and the noise model as it is in the parametric Box-Jenkins model.

Remark: It is clear that in a second step it should also be possible to use the knowledge from the estimated FRF $\hat{G}_{poly}$ to improve the generation of starting values for the numerical minimization of the cost function, but that is out of the scope of this paper.

III. Properties of the nonparametric Box-Jenkins model

In this section we discuss shortly the stochastic properties of the nonparametric Box-Jenkins estimate $\hat{\theta}_{np}$ (the minimizer of (13)). It can be shown that the consistency is not affected by replacing the parametric noise model by a nonparametric one, while there will be a small loss in efficiency with respect to the parametric noise model due to the larger number of parameters to be estimated in the nonparametric noise model. A detailed discussion is out of the scope of this paper, but the proofs follow the same lines as those given in [12]. The interested reader is referred to that paper or the book [4]. Here we briefly explain the major ideas.

A. Consistency

The consistency follows from the observation that the minimizer of the expected value of the cost function (13) is the same as that of the parametric Box-Jenkins model (4).

The calculation of the expected value $E\left\{\tilde{V}_{N,nonpar}\left(\theta\right)\right\}$ can be simplified because it can be shown that the numerator and denominator in (13) are asymptotically (for $N \to \infty$) independently distributed for Gaussian noise $v\left(t\right)$. The precise proof is out of the scope of this paper [15], but it is based on the observation that the sample mean and the sample variance of a Gaussian distributed stochastic are independent variables. Hence

$$E\left\{\tilde{V}_{N,nonpar}\left(\theta\right)\right\} = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} E\left\{\left|\tilde{E}\left(k\right)\right|^2\right\} E\left\{\frac{1}{\hat{\sigma}_v^2\left(k\right)}\right\}, \tag{14}$$

with $\tilde{E}\left(k\right) = \hat{Y}\left(k\right) - G\left(k,\theta\right)U_0\left(k\right)$. Define n_{dof} as the number of degrees of freedom in the χ^2-distributed variance estimate $\hat{\sigma}_v^2\left(k\right)$. The expected value $E\left\{1/\hat{\sigma}_v^2\left(k\right)\right\}$ is guaranteed to exist if $n_{dof} \geq 2$ (see Corollary 8.20 on page 304 of [4]). This is controlled by the default settings of the local polynomial method. The presence of a parameter independent bias in the variance estimate does not affect the consistency of $\hat{\theta}_{np}$, it only affects the efficiency. Moreover, this bias in (12) decreases towards zero for $N \to \infty$.

To proof the convergence of the cost function (13) to its expected value, $n_{dof} \geq 4$ is required because in this step 4^{th}order moments are used [12], [4].

B. Efficiency and distribution

The covariance matrix of $\hat{\theta}_{np}$ is shown to exist if $n_{dof} \geq 6$. Convergence towards a normal distribution is assured if $n_{dof} \geq 7$.

In [4] it is shown that there is a loss in efficiency with respect to the situation with an exactly known nonparametric noise model which is equal to

$$E_{loss} = \left(n_{dof} - 2\right) / \left(n_{dof} - 3\right). \tag{15}$$

C. Conclusion

Using the nonparametric noise model $\hat{\sigma}_v^2\left(k\right)$, obtained from the local polynomial method, in the nonparametric Box-Jenkins estimate results in consistent estimates that are asymptotically normal distributed. There is a slight loss in efficiency that depends on the number of degrees of freedom n_{dof}. This value is controlled by the default settings in the local polynomial method: the wider the local bandwidth used to fit the local polynomial, the larger n_{dof} will be, but at the same time the risk for bias errors grows. In this paper we used a default setting of $n_{dof} = 8$ in all the simulations.

IV. Comparison with earlier results

Nonparametric noise models were used for the first time in a Box-Jenkins identification approach in [14]. That paper starts mainly from the cost function in (13), making use of a parametric transient model, and a nonparametric noise model. The noise model is estimated from the residuals $Y\left(k\right) - G\left(k,\theta\right)U_0\left(k\right) - T\left(k,\theta\right)$ in (13) as explained in [14]. It resulted in noise estimates with errors of order $O\left(N^{-1}\right)$. So there are two major differences compared to the method that is presented in this paper: i) In this paper also the plant transient is eliminated in the nonparametric preprocessing step, while in [14] a parametric plant- and transient-model are estimated. In the initial step of [14], the transient is put equal to zero. Hence the initial noise model is also prone to the unmodeled transient errors during the initialization. ii) The errors of the nonparametric noise model that is obtained from the local polynomial method are smaller than those obtained with the method in [14]. For these reasons it is clear that the new method that is presented in this paper will be more robust (no parametric transient model needed, a better initial noise model is available to initialize the identification) and more efficient (a better nonparametric noise model is used).

V. Simulation results

Two simulations were defined: the first one identifies a 2^{nd} order system, while the second simulation deals with a 6^{th}order system. Three different noise filters are used: i) ARX-simulation $H_0 = 0.1/A_0$; ii) OE-simulation: $H_0 = 0.1$; iii) Box-Jenkins-simulation: $H_0 = 0.1 \times C_0/D_0$. The variance of the driving white noise is set equal to unity in all cases.

In each simulation we consider 100 realizations.

The excitation is filtered Gaussian noise, generated with the following filter coefficients: bGen=[0.5276 1.5829 1.5829 0.5276] and aGen=[1.0000 1.7600 1.1829 0.2781]. These can be calculated with the Matlab instruction: [bGen,aGen]=butter(3,2*fMax) with fMax=0.4. The zero mean driving white noise has unite variance.

The simulation length is 5500 points. The first 500 data points are not used in order to eliminate the initial transients of the simulation. This makes sure that the simulations do not always start with a system that is initially at rest.

Each data set is processed using the ARX-, OE-, Box-Jenkins-, and nonparametric Box-Jenkins method and for each of these the corresponding root mean square error is calculated as a function of the frequency. The simulations are done with a disturbing noise power level ranging from 1% to 30% of the undisturbed output power. The parameters of the local polynomial method were set such that n_{dof} =8. The results are discussed below.

A. 2^{nd}order system

The system G_0 is given by the filter coefficients b0=[0.1943 0.3885 0.1943] and a0 = [1.0000 0.7125 0.7449], generated with the Matlab instruction [b0,a0]=cheby1(2,10,2*fMax*0.9).

The noise in the Box-Jenkins-simulation is generated with c0=[0.1084 0.2169 0.1084], and d0=[1.0000 -0.8773 0.3111].

Figure 2 shows besides G_0 we also the power spectrum of the disturbing noise, and the $E_{RMS}(k) = rms\left(\hat{G}(k) - G_0(k)\right)$ for the different methods. From these results we learn that in each simulation the ARX, OE, and Box-Jenkins model are the best model for their respective simulation. This is expected from the theory, because these are the respective maximum likelihood solutions. However, it can also be seen that the nonparametric Box-Jenkins model is each time very close to this best result. This is again in agreement with the conclusions of the previous section. In this case the loss is well below 2 dB (20% on the standard deviation) which is in the order of magnitude given by (15) that predicts a loss of 10% on the standard deviation. This can be further reduced by increasing the bandwidth of the Local Polynomial method to increase n_{dof}.

This figure also shows that the use of wrong prior information can result in a very large increase of the rms-error. Especially the middle and right plot show a large increase of the error for the ARX-model with 40 dB or more (factor 100). Also the OE-model loses up to 10 dB in the right plot.

Conclusion: this simulation illustrates that the ARX and OE are optimal when the underlying assumptions are met. However these models are very sensitive to the validity of these assumptions. Box-Jenkins does very well in each of these simulations. It is very close (or equal) to the best solution. The nonparametric Box-Jenkins model is a robust alternative for the use of parametric noise models. In these three very different situations it produces results that are very close to the optimal solution, and this without any user interaction to tune the noise model to the specific situation.

B. 6^{th} order system

The settings in this simulation are similar to the previous ones, but the system is now given by b0 = [0.0040 0.0241 0.0604 0.0805 0.0604 0.0241 0.0040], and a0 =[1.0000 0.9468 1.7955 0.8304 1.7347 0.8808 0.9558], generated using the instruction [b0,a0]=cheby1(6,30,2*fMax*0.9).

The results are shown in Figure 3. We can draw the same conclusions as before, but here an additional aspect becomes visible: the OE and the parametric Box-Jenkins estimates suffer from local minima in this simulation while this is not the case for the nonparametric Box-Jenkins approach. The latter became now the best method in the OE and the Box-Jenkins simulations. Since the ARX-estimation is linear-in-the-parameters, is does not suffer from local minima, and it remains the best solution on the ARX-simulation, closely followed by the parametric- and nonparametric Box-Jenkins estimates.

C. Avoiding local minima

The previous section indicated that the parametric Box-Jenkins results were prone to local minima that result in poor estimates. This claim can be supported by verifying if the problem can be reduced by using improved initial estimates. In a first step, only the plant model is initialized using the results from the nonparametric Box-Jenkins model. The noise model was initialized at $H(q,\theta) = 1$. In a second step, also an improved initialization for the noise model was used: the residuals $e(t) = y(t) - G\left(q, \hat{\theta}_{npBJ}\right) u_0(t)$ are identified using an ARMA-model and these parameters are then used for the initialization of the noise model. The results in Figure 4 show that the loss due to initialization problems is already strongly reduced by the improved plant initialization. Only around the second resonance there is still a loss with respect to the nonparametric Box-Jenkins. When also the noise model is properly initialized, we retrieve the expected result: the parametric Box-Jenkins model is slightly better than the nonparametric Box-Jenkins result at all frequencies.

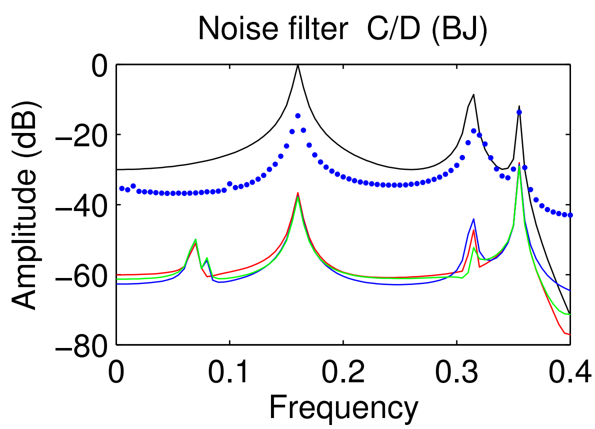

Figure 4. Avoiding local mininima in Box-Jenkins identification using improved starting values. Black line: G_0; red line: rms-error nonparametric Box-Jenkins; dotted blue: rms-error parametric Box-Jenkins; blue line: rms-error parametric Box-Jenkins, plant model initialized from the nonparametric method and initial noise model $H = 1$; green line: rms-error parametric Box-Jenkins, plant model initialized from the nonparametric Box-Jenkins, noise model: initialized from an ARMA-model fitted on the time domain residuals.

VI. Conclusions

We introduced a nonparametric Box-Jenkins model and compared it to ARX, OE, and Box-Jenkins identification. The major advantages of the nonparametric Box-Jenkins method are: i) its robustness to assumptions about the noise model structure and noise model order; ii) its user friendliness: no user decisions are required with respect to

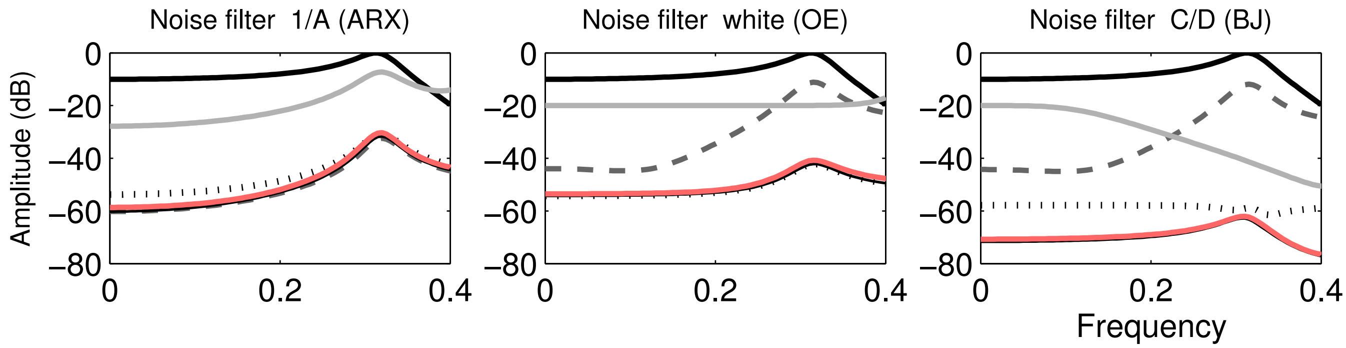

Figure 2. Comparison of ARX, OE, and Box-Jenkins with the nonparametric Box-Jenkins on a simple system. Bold line: G_0; Gray line: power spectrum disturbing noise; - -: rms-error ARX; rms-error ...: OE; —: rms-error Box-Jenkins; red line: rms-error nonparametric Box-Jenkins.

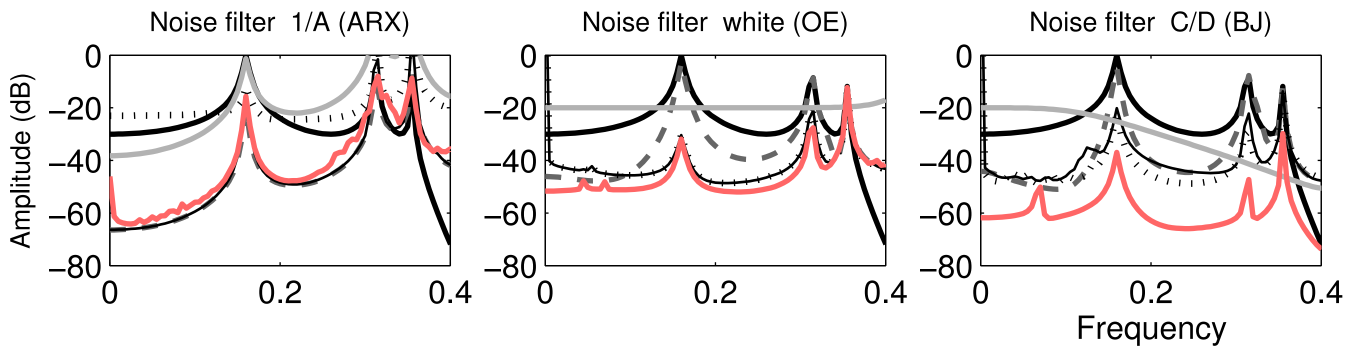

Figure 3. Comparison of ARX, OE, and Box-Jenkins with the nonparametric Box-Jenkins on a complex system. Bold line: G_0; Gray line: power spectrum disturbing noise; - -: rms-error ARX; ...: rms-error OE; —: rms-error Box-Jenkins; red line: rms-error nonparametric Box-Jenkins.

the noise model. The user is not requested to choose a noise model structure so that there is no risk of imposing wrong prior information. iii) Only the plant model order should be selected, there is no noise model order to be selected.

The nonparametric Box-Jenkins method turns out to be very close to the optimum for a wide range of situations. The price to be paid for this robust behavior is a slight loss in uncertainty of about 20% on the standard deviation. The study also reveals that the initialization process of the OE- and the Box-Jenkins model is critical, especially for more complex systems. Without a robust initialization procedure, these methods do not reach their global minimum and all optimal properties as obtained from the theory are lost.

References

[1] J.C. Agüero, J.I. Yuz, G.C. Goodwin, and R.A. Delgado, "On the equivalence of time and frequency domain maximum likelihood estimation," *Automatica* 46, 260-270, 2010.

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

[3] T. Söderström and P. Stoica, *System Identification.* Prentice-Hall, Englewood Cliffs 1989.

[4] R. Pintelon and J. Schoukens, *System Identification. A Frequency Domain Approach.* IEEE-press, Piscataway 2001.

[5] R. Pintelon, J.Schoukens, and G. Vandersteen, "Frequency domain system identification using arbitrary signals," *IEEE Trans. Automatic Control*, 42, 17171720, 1997.

[6] T. McKelvey, "Frequency domain identification methods," *Circuits Systems Signal Processing*, 21, 39-55, 2002.

[7] J. Schoukens, G. Vandersteen, K. Barbe, R. Pintelon, "Nonparametric Preprocessing in System Identification: a Powerful Tool," *European Journal of Control*, 2009; 15, pp. 260-274.

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

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

[10] J. Schoukens, Y. Rolain, and R. Pintelon, "Use of Parametric and Non-Parametric Noise-Models in Time and Frequency Domain System Identification," *49th IEEE Conference on Decision and Control,* 2010, Atlanta, USA, Paper WeA09.

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

[12] 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*, 1997; 33, pp. 1073-1086.

[13] K.R. Godfrey, *Perturbation signals for system identification*. Prentice Hall int., 1993.

[14] J. Schoukens, R. Pintelon, and Y. Rolain, "Box-Jenkins alike identification using nonparametric noise models," *Automatica,* 2004; 40, pp. 2083-2089.

[15] R. Pintelon, Proof of Lemma 12.2 in the 2nd edition of reference [4], to be published.