Incorporating noise modeling in dynamic networks using non-parametric models ⋆

Miguel Galrinho * **Niklas Everitt** * **Håkan Hjalmarsson** *

* *ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {galrinho,everitt,hjalmars}@ kth.se).*

Abstract: For identification of systems in dynamic networks, two-stage and instrumental variable methods are common time-domain methods. These methods provide consistent estimates of a chosen module of the network without estimating other parts of the network or noise models. However, disregarding noise modeling may come at a cost in estimation error. To capture the noise contribution, we propose the following procedure: first, we estimate a non-parametric model of an appropriate part of the network; second, we estimate the module of interest using signals simulated with the non-parametric model. The simulated signals are derived from an asymptotic maximum likelihood criterion. Preliminary simulations suggest that the propose method is competitive with existing approaches and is particularly beneficial with colored noise.

Keywords: System identification, networks, least-squares identification

1. INTRODUCTION

Dynamic networks are often too complex to be estimated as a whole using the prediction error method (Ljung, 1999). Estimating a chosen part of the network is computationally more appealing. However, using internal signals as inputs to the identification problem and applying a direct prediction error method (PEM) is often problematic: with feedback loops, the noise must be correctly modeled to obtain consistent estimates; with sensor noise, we have an errors-in-variables (EIV) setting, for which PEM is biased. Approaches to obtain consistent estimates of network modules have been studied by, among others, Dankers et al. (2016, 2015); Dankers and Van den Hof (2015); Gunes et al. (2014); Van den Hof et al. (2013).

In an EIV setting and without estimating a noise model, possible approaches include instrumental variable (IV) methods (Dankers et al., 2015) and two-stage methods (Van den Hof et al., 2013). In the latter, noiseless inputs are simulated from a non-parametric estimate of an appropriate part of the network. Although these methods are consistent, disregarding noise modeling may come at a cost in estimation error. Some frequency-domain methods deal with this issue: a fully non-parametric approach has been proposed by Dankers and Van den Hof (2015); the method proposed by Pintelon et al. (2010a,b) can also be used in this scenario, where a semi-parametric approach is taken—the noise model is captured by a non-parametric model and the plant model of interest is parametric.

In this paper, we propose a time-domain method to deal with the noise contribution also using semi-parametric approach. For single-input single-output (SISO) systems, this type of approach has been used by Everitt et al. (2016), named Model Order Reduction Steiglitz-McBride. Here, we extend this approach to estimate a module in a dynamic network. First, we estimate a non-parametric ARX model of an appropriate part of the network. Second, we estimate the module of interest using signals simulated from the ARX model and the Steiglitz-McBride method.

⋆ This work was supported by the Swedish Research Council under contracts 2015-05285 and 2016-06079.

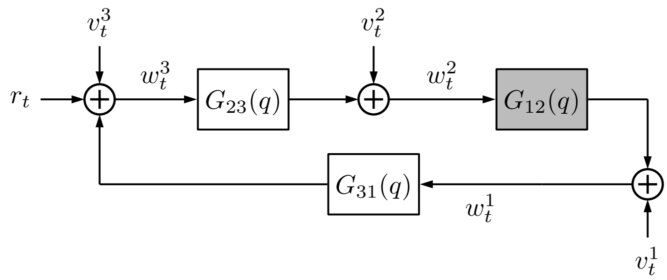

Fig. 1. Example of a dynamic network.

Although the basic idea resembles two-stage methods, the main contribution is that we use an asymptotic maximum likelihood (ML) criterion (Wahlberg, 1989) to derive the filters that simulate the signals. Because of the theoretical support on an ML criterion, we argue that the method may provide lower estimation error than alternative approaches. We support this argument by simulations, where we observe that the method is most beneficial for colored noise.

2. IDENTIFICATION OF SYSTEMS IN DYNAMIC NETWORKS: AN OVERVIEW

Consider the dynamic network in Fig. 1, written as

$$\begin{bmatrix} w_t^1 \\ w_t^2 \\ w_t^3 \end{bmatrix} = \begin{bmatrix} 0 & G_{12}(q) & 0 \\ 0 & 0 & G_{23}(q) \\ G_{31}(q) & 0 & 0 \end{bmatrix} \begin{bmatrix} w_t^1 \\ w_t^2 \\ w_t^3 \end{bmatrix} + \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ r_t \end{bmatrix}. \quad (1)$$

We have the following assumptions:

- $G_{12}(q)$, $G_{23}(q)$, and $G_{31}(q)$ are rational transfer functions in the delay operator q^{-1}, with at least one containing a delay;
- $\{v_t^1\}$, $\{v_t^2\}$, and $\{v_t^3\}$ are unknown process noise sequences given by Gaussian white noise sequences with finite variance filtered by unknown stable filters;
- $\{w_t^1\}$, $\{w_t^2\}$, and $\{w_t^3\}$ are measurable signals, whose measurements are given by
$$\tilde{w}_t^k = w_t^k + s_t^k, \tag{2}$$
$k = \{1, 2, 3\}$, where $\{s_t^1\}$, $\{s_t^2\}$, and $\{s_t^3\}$ are measurement noise sequences obtained by unknown stable filters driven by Gaussian white noise;
- $\{r_t\}$ is a known bounded reference signal, uncorrelated with the noise signals.

We will use this network throughout the paper to review available methods and to explain the proposed method. The objective is to estimate $G_{12}(q)$, parametrized as

$$G_{12}(q, \theta) = \frac{L_{12}(q, \theta)}{F_{12}(q, \theta)} = \frac{l_1 q^{-1} + \cdots + l_m q^{-m}}{1 + f_1 q^{-1} + \cdots + f_m q^{-m}}, \tag{3}$$

where $\theta = [f_1 \cdots f_m \; l_1 \cdots l_m]^\top$ are the parameters to estimate. We assume that there is unique $\theta = \theta_o$ such that $G_{12}(q) = G_{12}(q, \theta_o)$.

Alternatively to the recursive description in (1), the network can also be described by the relation from its external input signals to the measured output signals (2):

$$\begin{bmatrix} \tilde{w}_t^1 \\ \tilde{w}_t^2 \\ \tilde{w}_t^3 \end{bmatrix} = S(q) \begin{bmatrix} G_{12}(q)G_{23}(q) \\ G_{23}(q) \\ 1 \end{bmatrix} r_t + H_v(q) \begin{bmatrix} v_t^1 \\ v_t^2 \\ v_t^3 \end{bmatrix} + \begin{bmatrix} s_t^1 \\ s_t^2 \\ s_t^3 \end{bmatrix}, \tag{4}$$

with $S(q) = [1 - G_{12}(q)G_{23}(q)G_{31}(q)]^{-1}$ and

$$H_v(q) = S(q) \begin{bmatrix} 1 & G_{12}(q) & G_{12}(q)G_{23}(q) \\ G_{23}(q)G_{31}(q) & 1 & G_{23}(q) \\ G_{31}(q) & G_{12}(q)G_{31}(q) & 1 \end{bmatrix}.$$

Estimating the individual network modules with PEM using the complete network description can be problematic in several aspects. For this reason, identification in dynamic networks concerns estimating particular modules of interest. In our example, we are interested in $G_{12}(q)$. If the signals $\{w_t^1, w_t^2\}$ were known, the perhaps most natural approach to estimate $G_{12}(q)$ would be to use these signals as output and input to PEM. This is known as the direct approach (Van den Hof et al., 2013). Because there is feedback in the network, this approach has a disadvantage that the noise contribution $\{v_t^1\}$ must be correctly modeled to achieve a consistent estimate of $G_{12}(q)$. Nevertheless, the direct approach is not applicable in our setting: because $\{w_t^1, w_t^2\}$ are not known, but measured with noise according to (2), we need to use $\{\tilde{w}_t^1, \tilde{w}_t^2\}$ as output and input to the identification problem. This creates an errors-in-variables (EIV) problem, for which PEM provides biased estimates. We proceed to review two possible time-domain methods to consistently identify $G_{12}(q)$: the two-stage (or indirect) method and the instrumental variable (IV) method.

2.1 The Two-Stage Method

The two-stage method was proposed by Van den Hof and Schrama (1993) to obtain consistent estimates of a plant from closed-loop data without having to estimate a noise model. However, it can also be applied in the network case to simulate the desired input signal as follows (Van den Hof et al., 2013).

For the first stage, the measured input to $G_{12}(q)$ is (4)

$$\tilde{w}_t^2 = S(q)G_{23}(q)r_t + S(q)[G_{23}(q)G_{31}(q)v_t^1 + v_t^2 + G_{23}(q)v_t^3] + s_t^2.$$

Because $\{r_t\}$ is uncorrelated with the noise signals, estimating a high-order FIR model (without loss of generality, one delay is assumed in G_{23})

$$\tilde{w}_t^2 = \sum_{k=1}^{n} \eta_k q^{-k} r_t + e_t, \tag{5}$$

where n is the model order and $\{\eta_k\}_{k=1}^n$ are the parameters to estimate, provides a non-parametric estimate of $S(q)G_{23}(q)$ that is arbitrarily close to a consistent estimate as n increases. This stage provides estimates $\{\hat{\eta}_k\}_{k=1}^n$ of $\{\eta_k\}_{k=1}^n$, by solving the least-squares problem

$$\hat{\eta} = \left[\frac{1}{N}\sum_{t=1}^{N} \varphi_t^\top \varphi_t\right]^{-1} \left[\frac{1}{N}\sum_{t=1}^{N} \varphi_t \tilde{w}_t^2\right], \tag{6}$$

where N is the sample size, $\eta = [\eta_1 \cdots \eta_n]^\top$, and $\varphi_t = [r_{t-1} \cdots r_{t-n}]^\top$.

For the second stage, we simulate a noiseless input to $G_{12}(q)$ by

$$\hat{w}_t^2 := \sum_{t=1}^{n} \hat{\eta}_k q^{-k} r_t. \tag{7}$$

Then, PEM can be applied to the output-error (OE) model

$$\tilde{w}_t^1 = G_{12}(q, \theta)\hat{w}_t^2 + e_t, \tag{8}$$

which provides a consistent estimate of $G_{12}(q)$.

2.2 The Instrumental Variable Method

Instrumental variable methods (Söderström and Stoica, 1983) can provide consistent estimates of a plant without estimating a noise model and in an EIV setting (Thil et al., 2008). Thus, they are appropriate to estimate systems in dynamic networks (Dankers et al., 2015).

IV methods are a generalization of the least-squares method. To explain the idea, we start by using (1) and (2) to write

$$\tilde{w}_t^1 = \varphi_t^\top \theta + F_{12}(q)v_t^2 + F_{12}(q)s_t^1 - L_{12}(q)s_t^2,$$

where $\varphi_t = \begin{bmatrix} -\tilde{w}_{t-1}^1 \cdots -\tilde{w}_{t-m}^1 \; \tilde{w}_{t-1}^2 \cdots \tilde{w}_{t-m}^2 \end{bmatrix}^\top$. With an IV method, θ is estimated by solving

$$\hat{\theta} = \left[\frac{1}{N}\sum_{t=1}^{N} z_t^\top \varphi_t\right]^{-1} \left[\frac{1}{N}\sum_{t=1}^{N} z_t \tilde{w}_t^2\right], \tag{9}$$

where z_t is a vector containing an appropriate set of variables called instruments. If $z_t \equiv \varphi_t$, (9) reduces to the least-squares method, which provides biased estimates because the residual term

$$\tilde{w}_t^1 - \varphi_t^\top \theta = F_{12}(q)v_t^2 + F_{12}(q)s_t^1 - L_{12}(q)s_t^2 \tag{10}$$

is not white. Using (9) provides a consistent estimate if z_t is correlated with φ_t and uncorrelated with the residual term (10). Here, the reference signal $\{r_t\}$ may be used to construct the instruments because it is correlated with $\{\tilde{w}_t^1, \tilde{w}_t^1\}$ and uncorrelated with $\{v_t^2, s_t^1, s_t^2\}$. On the other hand, $\{\tilde{w}_t^3\}$ is not a candidate instrument because it is correlated with $\{v_t^2\}$. Although a generalized IV approach allows other instruments to be used (Dankers et al., 2015), we consider it outside the scope of this paper because it requires estimating a parametric noise model.

2.3 Summary and Potential Improvement

The reviewed approaches to estimate systems in dynamic networks focus on obtaining consistent estimates of particular modules, but they do not attempt to reduce the variance of the estimates. The latter problem has been considered, for example, in the two-stage method extension by Gunes et al. (2014). However, none of these approaches considers the noise signal properties when estimating a particular module. While we want to avoid estimating a parametric noise model, not taking the noise properties into account potentially increases the estimation error.

For SISO systems, there are approaches that capture the noise model with a non-parametric model and the plant with a parametric model: for example, the method by Schoukens et al. (2011) in the frequency domain and the method by Everitt et al. (2016) in the time domain. To estimate a module in a dynamic network with a similar semi-parametric approach, the frequency domain method by Pintelon et al. (2010a,b) can be applied. This method consists of two steps. In the first step, the parts of the noisy input and output signals to the module of interest that are correlated with the reference signal are estimated with the noise covariance, using the local polynomial method. In the second step, the non-parametric estimates of these signals and of the noise covariance are used as starting point to estimate a parametric plant model using the sample maximum likelihood (SML) estimator.

In this paper, we address the problem of noise modeling in dynamic networks in the time domain, by extending the Model Order Reduction Steiglitz-McBride (MORSM) method by Everitt et al. (2016) to this scenario. With this method, the parametric estimate of the plant is obtained from a non-parametric estimate, where the reduction is motivated by an asymptotic ML criterion. We proceed to review the SISO version of this method and the theoretical background that motivates it, which will be essential for the generalization to estimate modules in dynamic networks.

3. THE MODEL ORDER REDUCTION STEIGLITZ-MCBRIDE METHOD

Consider the SISO system

$$y_t = G(q)u_t + H(q)e_t, \tag{11}$$

where $G(q)$ and $H(q)$ can be described by rational transfer functions. We parametrize the plant by $G(q,\theta)$, similarly to (3). MORSM allows us to obtain an asymptotically efficient estimate of $G(q)$ without estimating a parametric model for $H(q)$ and without explicitly solving a non-convex optimization problem.

If PEM is used with data obtained in open loop, $H(q)$ can be over-parametrized without affecting the asymptotic properties of the estimated $G(q,\theta)$. However, choosing the noise-model order arbitrarily large will make the problem computationally more difficult for PEM. The exception is if an ARX model structure is used, in which case the global minimum of the prediction error criterion can be found by least squares. This consists in estimating the model

$$A(q,\eta)y_t = B(q,\eta)u_t + e_t, \tag{12}$$

where

$$A(q,\eta) = 1 + \sum_{k=1}^{n} a_k q^{-k}, \quad B(q,\eta) = \sum_{k=1}^{n} b_k q^{-k},$$

and $\eta = [a_1 \; \cdots \; a_n \; b_1 \; \cdots \; b_n]^\top$, providing estimates $\hat{\eta}$.

If the ARX-model order n is allowed to be arbitrarily large, (12) models (11) with arbitrary accuracy (Ljung and Wahlberg, 1992). Then, asymptotically in model order, the ARX-model estimate can be used to obtain non-parametric estimates of $G(q)$ and $H(q)$ by

$$\hat{G}(q) = \frac{B(q,\hat{\eta})}{A(q,\hat{\eta})}, \quad \hat{H}(q) = \frac{1}{A(q,\hat{\eta})}. \tag{13}$$

The asymptotic distribution of the non-parametric estimate $G(q,\hat{\eta})$ can be characterized in the frequency domain by (Wahlberg, 1989)

$$\hat{G}(e^{j\omega}) - G(e^{j\omega}) \sim \mathcal{N}\left(0, \frac{n}{N}\frac{\Phi_v(e^{i\omega})}{\Phi_u(e^{i\omega})}\right),$$

where $\mathcal{N}$ stands for the normal distribution, $\Phi_u(e^{i\omega})$ is the spectrum of the input and $\Phi_v(e^{i\omega})$ is the spectrum of the noise term $v := H(q)e_t$. This motivates that the reduction from the non-parametric estimate $\hat{G}(q)$ to a parametric one—$G(q,\theta)$—be performed by minimizing the asymptotic maximum likelihood criterion

$$V_N(\theta) = \int_0^{2\pi} \left|\hat{G}(e^{i\omega}) - G(e^{i\omega},\theta)\right|^2 \frac{\Phi_u(e^{i\omega})}{\Phi_v(e^{i\omega})} d\omega.$$

Because $\Phi_v(e^{i\omega})$ is typically unknown, we need to replace it by an estimate. As shown by Wahlberg (1989), $\Phi_v(e^{i\omega})$ may be replaced by its non-parametric (scaled) estimate $|\hat{H}(q)|^2$ without affecting the asymptotic properties of the estimate of $G(q,\theta)$. Thus, minimizing

$$V_N(\theta) = \int_0^{2\pi} \left|\hat{G}(e^{i\omega}) - G(e^{i\omega},\theta)\right|^2 \frac{\Phi_u(e^{i\omega})}{|\hat{H}(e^{i\omega})|^2} d\omega \tag{14}$$

provides an asymptotic efficient estimate of θ.

In the time domain, a consistent estimate of (14) for finite sample size is given by

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{\hat{G}(q) - G(q,\theta)}{\hat{H}(q)}u_t\right]^2. \tag{15}$$

In turn, using (13), we can re-write (15) as

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[B(q,\hat{\eta})u_t - G(q,\theta)A(q,\hat{\eta})u_t\right]^2. \tag{16}$$

The ASYM method (Zhu, 2001) consists in minimizing (16), which is an OE problem with a simulated output and filtered input, defined by

$$\hat{y}_t := B(q,\hat{\eta})u_t, \quad \hat{u}_t := A(q,\hat{\eta})u_t. \tag{17}$$

Using PEM, a non-convex optimization routine is required to minimize (16). An alternative is to use the Steiglitz-McBride method (Stoica and Söderström, 1981), which uses least squares iteratively (we refer to it as *Steiglitz-McBride*). This method is not asymptotically efficient when applied to the measured data (Stoica and Söderström, 1981); however, when applied to the simulated data set (17), Steiglitz-McBride provides asymptotically efficient estimates in one iteration (Everitt et al., 2016).

Motivated by this, the Model Order Reduction Steiglitz-McBride method consists of the following two steps: first, estimate a high-order ARX model, using least squares; second, apply Steiglitz-McBride to the data set (17), which is motivated by an asymptotic ML criterion.

We identify three advantages compared to a direct PEM estimation. First, we transform a Box-Jenkins problem into an OE one, which is computationally simpler. Second, the user does not need to choose a parametrization for the noise model. Third, we avoid a non-convex optimization routine by applying Steiglitz-McBride. These properties make MORSM an appealing method to estimate systems in dynamic networks.

4. MORSM FOR DYNAMIC NETWORKS

To generalize MORSM to dynamic networks, consider the single-input multi-output (SIMO) ARX model from r_t to the input and output signals of $G_{12}(q)$ (the module of interest), w_t^2 and w_t^1, respectively:

$$A(q,\eta)\begin{bmatrix} w_t^2 \\ w_t^1 \end{bmatrix} = B(q,\eta) r_t + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}$$

where $A(q,\eta)$ is 2×2 , $B(q,\eta)$ is 2×1, and the parameter vector η contains all the polynomials in these transfer matrices. An estimate $\hat{\eta}$ can then be obtained by least squares, which allows us to obtain non-parametric estimates

$$\begin{bmatrix} \hat{T}_{23}(q) \\ \hat{T}_{13}(q) \end{bmatrix} = A^{-1}(q,\hat{\eta}) B(q,\hat{\eta}), \tag{18}$$

where

$$\begin{bmatrix} T_{23}(q) \\ T_{13}(q) \end{bmatrix} := S(q) \begin{bmatrix} G_{23}(q) \\ G_{12}(q) G_{23}(q) \end{bmatrix}.$$

Having estimates of $T_{13}(q)$ and $T_{23}(q)$, the asymptotic ML estimate of $G_{12}(q)$ can be derived from these estimates, using that they are related by $T_{13}(q) - G_{12}(q) T_{23}(q) = 0$. To do this, we start by writing the asymptotic distribution

$$\hat{T}_{13}(e^{j\omega}) - G_{12}(e^{j\omega}) \hat{T}_{23}(e^{j\omega}) \sim \mathcal{N}\left(0, \frac{n}{N} \frac{P(e^{j\omega},\theta_o)}{\Phi_r(e^{j\omega})}\right), \tag{19}$$

where $\Phi_r(e^{j\omega})$ the spectrum of the reference signal r_t and

$$P(e^{j\omega},\theta) = \begin{bmatrix} -G_{12}(e^{j\omega},\theta) & 1 \end{bmatrix} \Phi_{\bar{v}}(e^{i\omega}) \begin{bmatrix} -G_{12}(e^{-j\omega},\theta) \\ 1 \end{bmatrix}, \tag{20}$$

where $\Phi_{\bar{v}}(e^{i\omega})$ is the spectrum of the noise signal

$$\bar{v}_t := \begin{bmatrix} \tilde{w}_t^2 \\ \tilde{w}_t^1 \end{bmatrix} - \begin{bmatrix} T_{23}(q) \\ T_{13}(q) \end{bmatrix} r_t. \tag{21}$$

This suggests that the model reduction step should be according to the cost function

$$V_N(\theta) = \int_0^{2\pi} |\hat{T}_{13}(e^{j\omega}) - G_{12}(e^{j\omega},\theta) \hat{T}_{23}(e^{j\omega})|^2 \frac{\Phi_r(e^{j\omega})}{P(e^{j\omega},\theta_o)} d\omega. \tag{22}$$

As with MORSM, we replace $\Phi_{\bar{v}}(e^{j\omega})$ in (20) by a non-parametric estimate—that is, the spectrum of the signal

$$\hat{\bar{v}}_t = \begin{bmatrix} \tilde{w}_t^2 \\ \tilde{w}_t^1 \end{bmatrix} - \begin{bmatrix} \hat{T}_{23}(q) \\ \hat{T}_{13}(q) \end{bmatrix} r_t \tag{23}$$

instead of (21)—and apply Steiglitz-McBride to perform the model reduction. However, we have two problems to address. First, the term $P(e^{j\omega},\theta_o)$ depends on the true parameters θ_o. Second, a spectral factor of this term is not available in closed form to simulate the signals.

Concerning the first problem, we replace θ_o by a consistent estimate. Because it is an ML criterion, this can be done without affecting the asymptotic properties of the ML estimate (Wahlberg, 1989). To obtain a consistent estimate, we may set $P(e^{j\omega},\theta_o)$ in (22) to one and minimize

$$V_N^0(\theta) = \frac{1}{N}\sum_{t=1}^{N} \left[\left(\hat{T}_{13}(e^{j\omega}) - G_{12}(e^{j\omega},\theta)\hat{T}_{23}(e^{j\omega})\right) r_t\right]^2.$$

To do this, we apply Steiglitz-McBride with the data set

$$\hat{y}_t^0 = \hat{T}_{13}(q) r_t, \quad \hat{u}_t^0 = \hat{T}_{23}(q) r_t, \tag{24}$$

which provides a consistent estimate $\hat{\theta}$ that we use to replace θ_o in (22).

Concerning the second problem, a standard approach to obtain a spectral factorization of $P(e^{j\omega},\hat{\theta})$ is to write the state-space form of the signal $\begin{bmatrix} G_{12}(q,\hat{\theta}) & 1 \end{bmatrix} \hat{\bar{v}}_t$ and compute the Kalman filter. However, numerical problems might occur when solving the Riccati equation because of the potential high order of the model. Because we only need a non-parametric estimate of the spectral factor, we instead fit an AR model $D(q)\hat{x}_t = e_t$ using least squares, where

$$D(q) = 1 + \textstyle\sum_{k=1}^{n} d_k q^{-k}, \quad \hat{x}_t := \begin{bmatrix} -G_{12}(q,\hat{\theta}) & 1 \end{bmatrix} \hat{\bar{v}}_t. \tag{25}$$

Then, if $\hat{D}(q)$ is the AR-model estimate, we have that $|\hat{D}(q)|^2 \approx P^{-1}(e^{j\omega},\hat{\theta})$. Finally, minimizing (22) with this estimate of $P^{-1}(e^{j\omega},\hat{\theta})$ corresponds to minimizing

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \left[\left(\hat{T}_{13}(e^{j\omega}) - G_{12}(e^{j\omega},\theta)\hat{T}_{23}(e^{j\omega})\right)\hat{D}(q) r_t\right]^2,$$

which provides the final data set for the Steiglitz-McBride:

$$\hat{y}_t := \hat{T}_{13}(q)\hat{D}(q) r_t, \qquad \hat{u}_t := \hat{T}_{23}(q)\hat{D}(q) r_t. \tag{26}$$

In summary, the MORSM algorithm for networks consists of the following steps:

(1) estimate an ARX model and construct estimates $\hat{T}_{13}(q)$ and $\hat{T}_{23}(q)$, according to (18);
(2) apply the Steglitz McBride method to the simulated data set (24), providing a consistent estimate $\hat{\theta}$;
(3) estimate an AR model from the signal $\hat{x}_t$, defined by (25) and (23);
(4) apply the Steglitz-McBride method to the simulated data set (26), with $\hat{D}(q)$ obtained in Step 3.

5. SIMULATION STUDY

In this section, we perform a simulation study covering three cases with different noise signal spectra. For all cases, the transfer function we are interested in estimating is

$$G_{12}(q) = \frac{q^{-1}}{1-0.8q^{-1}},$$

while the remaining modules in the network are

$$G_{23}(q) = \frac{0.2q^{-1}}{1-0.5q^{-1}}, \quad G_{31}(q) = \frac{-0.1q^{-1}}{1-0.4q^{-1}}.$$

We generate the reference signal by $r_t = [1 - 0.7q^{-1}]^{-1} e_t^r$, where $\{e_t^r\}$ is Gaussian white noise with unit variance. The sample size is $N = 5000$.

We compare the following methods:

- two-stage method (2-Stage), estimating an FIR model of order $n = 30$ (first stage) according to (5) and (6), and estimating an OE model using PEM (second stage) with (8) and simulated input (7);
- instrumental variable method (IV), estimating (9) with $z_t = [r_{t-3}\ r_{t-4}\ r_{t-5}]^\top$;
- the sample maximum likelihood method (SML), according to the implementation in the MATLAB toolbox (version October 2011) complementing the book by Pintelon and Schoukens (2012), with 50 degrees of freedom for the covariance estimate and order 1 for the local polynomial approximation;
- the proposed method (MORSM), estimating an ARX model of order $n = 15$ (first step), an AR model of order $n = 30$ (third step), and using 5 Steiglitz-McBride iterations (second and fourth steps).

For SML, the MIMO local polynomial method is first used with $\{r_t\}$ as input and $\{\tilde{w}_t^1, \tilde{w}_t^2\}$ as outputs; then, the obtained non-parametric frequency-domain estimates of the latter signals and of the noise covariance are used with the sample maximum likelihood method. We obtain the initial estimate for the optimization problem by applying iterative quadratic maximum likelihood (IQML), which, in turn, is initialized by weighted total least squares (WTLS)—functions that are available in the toolbox. A maximum number of 100 iterations (default) is used.

The settings for all the methods were chosen based on a few empirical observations regarding performance. In a more extensive simulation study, a data-based selection of these settings should be used instead. However, we do not consider it for the purpose of this preliminary simulation.

For illustrative purposes, it would be interesting to observe how much is gained from Step 2 to Step 4 of the proposed method, as well as how much is lost by not using θ_o when obtaining the spectral factor. With this purpose, the following estimates are also included for comparison:

- Step 2 of the proposed method (naive MORSM);
- the proposed method with $G_{12}(q, \theta_o)$ used in Step 3 when constructing (25) (oracle MORSM).

We evaluate the performance using the FIT of the estimated impulse response of $G_{12}(q)$, given in percent by

$$\text{FIT} = 100 \left(1 - \frac{||g^{12} - \hat{g}^{12}||_2}{||g^{12} - \text{mean}(g^{12})||_2} \right),$$

where g^{12} and $\hat{g}^{12}$ are vectors with the impulse coefficients of $G_{12}(q)$ and $G_{12}(q, \hat{\theta})$ (for a particular method), respectively. We perform 100 Monte Carlo runs.

The three cases we study are:

I) the process and sensor noises are uncorrelated Gaussian white noise sequences with unit variance;

II) the process noise sequences are low-pass signals generated by $v_t^k = H(q) e_t^{v_k}$, $k = \{1, 2, 3\}$, where $H(q) = [1 - 0.95q^{-1}]^{-1}$, and $\{e_t^{v_k}\}$ are mutually uncorrelated white Gaussian noise sequences with unit variance, and the sensor noise sequences are as in Case I).

III) the process noise sequences are as in Case II), and the sensor noise sequences are generated to have the same spectra, by $s_t^k = H(q) e_t^{s_k}$, $k = \{1, 2, 3\}$, where $\{e_t^{s_k}\}$ are mutually uncorrelated white Gaussian noise sequences with unit variance.

Although correlation between signals should not affect the methods, we consider the uncorrelated case for simplicity.

Case I) The results for the case where all noise contributions are white are presented in the left plot of Fig. 2. In this case, there is hardly any improvement from Step 2 of the proposed method (naive MORSM) to the complete method (MORSM), which also has little advantage over the two-stage method (2-Stage). The reason is that the total noise contribution when driven by white noise signals does not have a considerable influence, and may be disregarded at almost no cost in performance. Moreover, there is no observable difference between the proposed method and using knowledge of the true system to filter the signals in Step 3 (oracle MORSM), meaning that our proposal of replacing the true system in Step 3 by the estimate obtained in Step 2 is acceptable in this case. The IV method we use is not competitive with the remaining methods here. Finally, SML and MORSM have similar performance.

Case II) The results for the case where the process noise signals are low pass are presented in the middle plot of Fig. 2. In this case, there is a clear improvement from naive MORSM to MORSM, while the 2-Stage is not competitive. Thus, when the noise contributions are sufficiently correlated in time, performing the filtering derived from the asymptotic ML approach is beneficial. Moreover, replacing the true system with a consistent estimate in the ML weighting does not deteriorate our approach here either, as MORSM and the oracle version have identical performance. The IV method we use has better performance than in the previous case, because it benefits from additional correlations in the internal network signals. Like MORSM, the SML method is also capable of handling the colored noise signals used.

Case III) The results for the case where both the process and sensor noise signals are low-pass signals are presented in the right plot of Fig. 2. In this case, the estimate naive MORSM is very poor, but the improvement to the complete method is considerable. Also here, MORSM performs similarly to the case that the true system is used in Step 3 (oracle MORSM), even if the true system is replaced by a poor estimate (the one obtained by naive MORSM). The SML and MORSM are competitive in this simulation, but 2-Stage and IV are not.

In summary, our simulations suggest that the proposed method can provide a smaller estimation error than standard two-stage and IV methods. If the noise contributions are not sufficiently colored, the two-stage method performs very similar to the proposed method. With increasingly colored process noise, IV methods may benefit from additional correlations in the internal network signals, but performance of the two-stage method deteriorates. If also the sensor noise is highly colored, performance of both IV and two-stage is affected. Meanwhile, the proposed

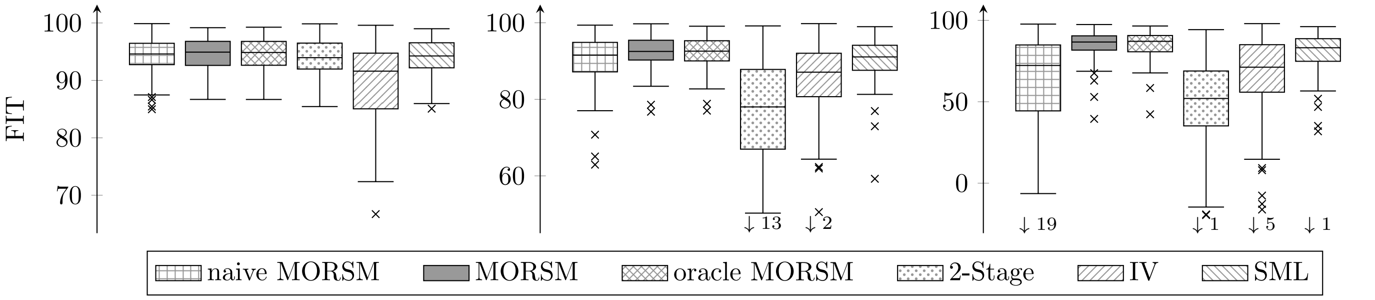

Fig. 2. Simulation studies with three cases: Case I) white process and sensor noise signals (left); Case II) low-pass process noise signals and white sensor noise signals (middle); Case III) low-pass process and sensor noise signals.

method showed good performance in all the cases studied, and is competitive with the SML method. More extensive simulation studies are required to test the robustness and limitations of the method.

6. CONCLUSION

In this paper, we proposed a method for estimation of systems in dynamic networks. The method resembles standard two-stage methods, but we motivate the simulated signals using asymptotic ML. We argue that this should decrease the estimation error compared with two-stage and IV methods, and support this argument with simulations.

The proposed method showed also competitive performance with the frequency-domain sample maximum likelihood method. These methods are conceptually similar, although using different approaches. Besides the time- and frequency-domain difference, MORSM uses a non-parametric ARX model to capture the dynamics of the system, while SML uses the local polynomial method to obtain a non-parametric estimate of the noise and the frequency response function. In both cases, the parametric estimate is motivated by an ML criterion. A more in-depth comparison between these methods—both theoretical and experimental—is considered for future work.

To keep the notation simple, we used a specific network for our discussion. This network has only one external reference and the output signal we are interested in is generated only by the module of interest (i.e., the row of the transfer matrix in (1) containing the module of interest has no other non-zero entries). If this is not the case, (19) will be multivariate. Because multivariate asymptotic ML is covered by Zhu (2001), we can use the results therein to generalize the proposed method to cover such cases. We will address this in future work.

REFERENCES

Dankers, A. and Van den Hof, P. (2015). Non-parametric identification in dynamic networks. In *54th Conference on Decision and Control*, 895–900. Osaka, Japan.

Dankers, A., Van den Hof, P., Bombois, X., and Heuberger, P. (2015). Errors-in-variables identification in dynamic networks—consistency results for an instrumental variable approach. *Automatica*, 62, 39–50.

Dankers, A., Van den Hof, P.M., Bombois, X., and Heuberger, P.S. (2016). Identification of dynamic models in complex networks with prediction error methods: Predictor input selection. *IEEE Transactions on Automatic Control*, 61(4), 937–952.

Everitt, N., Galrinho, M., and Hjalmarsson, H. (2016). Optimal model order reduction with the Steiglitz-McBride method. *submitted to Automatica (arXiv:1610.08534)*.

Gunes, B., Dankers, A., and Van den Hof, P. (2014). Variance reduction for identification in dynamic networks. In *19th IFAC World Congress*, 2842–2847.

Ljung, L. (1999). *System Identification. Theory for the User*. Prentice-Hall.

Ljung, L. and Wahlberg, B. (1992). Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Advances in Applied Probabilities*, 24, 412–440.

Pintelon, R., Schoukens, J., Vandersteen, G., and Barb, K. (2010a). Estimation of nonparametric noise and FRF models for multivariable systems—Part I: theory. *Mechanical Systems and Signal Processing*, 24(3), 573–595.

Pintelon, R., Schoukens, J., Vandersteen, G., and Barb, K. (2010b). Estimation of nonparametric noise and FRF models for multivariable systems—Part II: extensions, applications. *Mechanical Systems and Signal Processing*, 24(3), 596–616.

Pintelon, R. and Schoukens, J. (2012). *System Identification: A Frequency Domain Approach*. Wiley.

Schoukens, J., Rolain, Y., Vandersteen, G., and Pintelon, R. (2011). User friendly Box-Jenkins identification using nonparametric noise models. In *50th IEEE Conference on Decision and Control*. Orlando, Florida, USA.

Söderström, T. and Stoica, P. (1983). *Instrumental Variable Methods for System Identification*. Springer.

Stoica, P. and Söderström, T. (1981). The Steiglitz-McBride identification algorithm revisited—convergence analysis and accuracy aspects. *IEEE Transactions on Automatic Control*, 26(3), 712–717.

Thil, S., Gilson, M., and Garnier, H. (2008). On instrumental variable-based methods for errors-in-variables model identification. In *17th IFAC World Congress*, 426–431.

Van den Hof, P., Dankers, A., Heuberger, P., and Bombois, X. (2013). Identification of dynamic models in complex networks with prediction error methods—Basic methods for consistent module estimates. *Automatica*, 49(10), 2994–3006.

Van den Hof, P. and Schrama, R. (1993). An indirect method for transfer function estimation from closed loop data. *Automatica*, 29(6), 1523–1527.

Wahlberg, B. (1989). Model reduction of high-order estimated models: the asymptotic ML approach. *International Journal of Control*, 49(1), 169–192.

Zhu, Y. (2001). *Multivariable System Identification for Process Control*. Pergamon.