2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

Non-Parametric Identification in Dynamic Networks

Arne Dankers and Paul M.J. Van den Hof

***Abstract*— In this paper we present a non-parametric approach to identification in networks. The main advantage of a non-parametric approach is that consistent estimates can be obtained with very little prior knowledge about the system. This is a particularly important consideration for a network identification problem which can easily become very complex with high order dynamics and many inputs. We consider a very general framework for dynamic networks that includes measured variables, external excitation variables, process noise, and sensor noise.**

I. Introduction

Identification in dynamic networks is an increasingly important area of system identification as systems become larger and more complex. Many systems in engineering can be modeled as dynamic networks: power systems, communications systems, pipelines, etc. In the literature there is increasing activity in network identification tackling various challenges such as investigating under which conditions consistent estimates are possible [1], variance issues related to networks [2], [3], numerical issues [4], and issues that arise when the network structure is unknown [5], [6], [7]. Interestingly, identification in networks offers many advantages to classical open and closed loop identification: sensor noise is easily dealt with [8], and the variance of the resulting estimates can be reduced by using additional measurements obtained from the network [2].

In this paper we take a non-parametric approach to identification in dynamic networks. Non-parametric methods are part of the standard toolbox of system identification [9], [10]. The advantage of using a non-parametric approach is that it offers an expression for the system dynamics with very little prior information. In particular there is no need to specify the number of poles and zeros of each module a priori. This is a significant advantage, especially when a large number of modules need to be estimated. Another major advantage is that when periodic excitation is used, consistent estimates of the transfer function of interest are possible with only very weak assumptions on the noise. This is important in a network setting because many of the noise variables are likely to be correlated. The main disadvantage of the non-parametric approaches is that the variance of the estimates can be relatively high [9].

In this paper we consider a very general framework for a dynamic network consisting of *internal variables* that are measurable and dynamically interrelated, *external variables* that are directly manipulable by the user, *process noise* that is a random stochastic process affecting the value of the internal variables, and *sensor noise* that is a random stochastic error in the recording of the value of an internal variable. In this paper we do not assume that each internal variable is measured. The ideas presented in this paper are extensions of the idea of *predictor input selection* as presented in [11] and the errors-in-variables methods presented in [8] from a parametric to a non-parametric framework. The motivation for this extension is, as described above, to increase the practicality of identification in dynamic networks.

In Section II we present background material on dynamic networks. In Section III we present the main result split into three different cases: periodic external excitation present, arbitrary excitation present, and a method based on using non-parametric noise models to achieve consistency.

A. Dankers is with the Electrical and Computer Engineering Dept. at the University of Calgary, Canada adankers@hifieng.com

P. M. J. Van den Hof is with the Dept. of Electrical Engineering, Eindhoven University of Technology, The Netherlands p.m.j.vandenhof@tue.nl

II. Dynamic Networks

The networks that are considered in this paper are built up of L elements (or nodes), related to L scalar *internal variables* w_j, $j = 1, \ldots, L$ [1]. It is assumed that each internal variable is such that it can be written as:

$$w_j(t) = \sum_{k \in \mathcal{N}_j} G^0_{jk}(q) w_k(t) + r_j(t) + v_j(t) \qquad (1)$$

where $G^0_{jk}(q)$, $k \in \mathcal{N}_j$ is a proper rational transfer function, q^{-1} is the delay operator (i.e. $q^{-1}u(t) = u(t-1)$) and,

- $\mathcal{N}_j$ is the set of indices of internal variables with direct causal connections to w_j, i.e. $i \in \mathcal{N}_j$ iff $G^0_{ji} \neq 0$;
- v_j is an unmeasured *process disturbance variable* modelled as a stationary stochastic process with rational spectral density, i.e. $v_j = H^0_j(q)e_j$ where e_j is a white noise process, and H^0_j is a monic, stable, minimum phase transfer function;
- r_j is an *external variable* that is known and can be manipulated by the user; it is an important variable that can provide user-chosen excitation to the network.

Throughout this paper superscript 0 will be used to denote the dynamics of the data generating system. It may be that the disturbance and/or external variable are not present at some nodes. The entire network is defined by:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G^0_{12} & \cdots & G^0_{1L} \\ G^0_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G^0_{L-1\ L} \\ G^0_{L1} & \cdots & G^0_{L\ L-1} & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_L \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_L \end{bmatrix},$$

978-1-4799-7885-4/15/$31.00 ©2015 IEEE

Using an obvious notation this results in:

$$w = G^0 w + r + v \tag{2}$$

where, w, r, and v are vectors. If an external or noise variable is absent at node i, the ith entry of r or v respectively is 0.

Internal variables are measured using sensors. Since errors are made when recording the value of a variable the measured version of an internal variable w_k is not the same as the actual value of the internal variable. The measurement of w_k is denoted $\tilde{w}_k$, and the error in the measurement is called *sensor noise*:

$$\tilde{w}_k = w_k + s_k \tag{3}$$

where s_k is the sensor noise and is modelled as a stochastic process with rational power spectral density. Equations (2) and (3) together define the *data generating system*.

Often in a network, process noise variables are correlated. For instance, in a mechanical structure buffeting wind may cause disturbances in the measured positions of joints in the structure. It is likely that the noise variables affecting different joints in the structure will be correlated since they are caused by the same wind.

All networks considered in this paper satisfy the following general assumption to ensure stability and causality.

Assumption 1:

(a) The network is well-posed in the sense that all principal minors of $\lim_{z\to\infty}(I - G^0(z))$ are non-zero.

(b) $(I - G^0)^{-1}$ is stable.

(c) All r_m, $m \in \mathcal{R}$ are uncorrelated to all v_k, $k \in \mathcal{V}$.[1]

Cross correlation and cross power spectral densities play an important role in this paper. They are defined as follows. The auto and cross correlation of vectors of variables x and y are defined as

$$R_x(\tau) = \bar{\mathbb{E}}[x(t)x^T(t-\tau)], \quad R_{xy}(\tau) = \bar{\mathbb{E}}[x(t)y^T(t-\tau)] \tag{4}$$

respectively, where $\bar{\mathbb{E}}[\ \cdot\ ] = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}[\ \cdot\ ]$, and $\mathbb{E}$ denotes the expected value operator. The power spectral density and cross power spectral density are

$$\Phi_x(\omega) := \mathcal{F}[R_x(\tau)] \text{ and } \Phi_{xy}(\omega) := \mathcal{F}[R_{xy}(\tau)]$$

respectively, where $\mathcal{F}[\ \cdot\ ]$ denotes the Fourier Transform.

III. Non-Parametric System Identification in Dynamic Networks

We present three different methods for non-parametrically obtaining estimates of G^0_{ji} embedded in a dynamic network: a method based on periodic external inputs; a method for arbitrary external inputs; and a method based on non-parametric noise modeling.

The advantage of using periodic excitation is that only weak assumptions need to be made about the process and sensor noise. This is an important consideration for identification in networks because it is likely that many of the process noise variables are correlated, and the internal variables are measured with sensor noise (that may be correlated).

We also present an analogous method but extended to arbitrary external excitation. The price for allowing any type of external excitation is that stronger assumptions on the noise are required.

The third method is based on using non-parametric noise modeling. For this method no external excitation is required as long as there are measured internal variables available other than those selected as predictor inputs.

To obtain an estimate, first predictor inputs must be selected (these are the variables that are used to predict the output w_j). Rules for choosing predictor inputs such that consistent estimates of G^0_{ji} are possible are presented in [11]. We briefly summarize these rules in the following section.

A. Predictor Input Selection

When attempting to predict the value of a variable w_j in a dynamic network, it is important to properly choose which variables are used to predict w_j. If the predictor inputs are not chosen properly, then consistent estimates of G^0_{ji} are not possible. The set of internal variables chosen as predictor inputs is denoted as $w_k, k \in \mathcal{D}_j$. The set of external variables chosen as predictor inputs is denoted $r_k, k \in \mathcal{P}_j$. From [11] the rules for choosing $\mathcal{D}_j$ is as follows.

Property 1: Consider the set of predictor inputs w_ℓ, $\ell \in \mathcal{D}_j$, and the "output" w_j. Let $\mathcal{D}_j$ satisfy the conditions:

(a) $i \in \mathcal{D}_j$, $j \notin \mathcal{D}_j$,

(b) every loop from w_j to w_j passes through a w_k, $k \in \mathcal{D}_j$,

(c) every path from w_i to w_j, excluding the path G^0_{ji}, passes through a w_k, $k \in \mathcal{D}_j$. □

In the sequel it will be very useful to express w_j (the output of the transfer function of interest) in terms of the variables selected as predictor inputs w_k, $k \in \mathcal{D}_j$ and r_k, $k \in \mathcal{P}_j$). Consider the following proposition.

Proposition 1 ([11]): Consider a dynamic network as defined in Section II that satisfies Assumption 1, and a set of predictor inputs $\mathcal{D}_j$. Then $w_j(t)$ can be uniquely written as

$$w_j(t) = \sum_{k\in\mathcal{D}_j} \breve{G}^0_{jk}(q) w_k(t) + \sum_{k\in\mathcal{Z}_j} \breve{F}^0_{jk}(q)\big(r_k(t) + v_k(t)\big) + \\ + r_j(t) + v_j(t) \tag{5}$$

where $\mathcal{Z}_j = \{1, \ldots, L\} \setminus \{\mathcal{D}_j \cup \{j\}\}$, and $\breve{G}^0_{ji}(q) = G^0_{ji}(q)$ if $\mathcal{D}_j$ has Property 1. □

The proposition is proved in [11]. Note that (unlike in (2)) in (5) there is a transfer function $\breve{F}^0_{jk}$ present associated with the external variables. It is convenient to lump all the process noise terms in (5) together as:

$$\breve{v}_j = \sum_{k\in\mathcal{Z}_j} \breve{F}^0_{jk}(q) v_k(t) + v_j(t) = \breve{H}^0_j(q)\breve{e}_j(t). \tag{6}$$

where $\breve{H}_j$ is a (unique) monic, stable, minimum phase transfer function, and $\breve{e}_j$ is white noise.

[1]Throughout this paper r uncorrelated to v will mean that the cross-correlation function $R_{rv}(\tau)$ is zero for all τ.

Proposition 1 provides a rule for choosing which internal variables to choose as predictor inputs given the objective of identifying G^0_{ji} (i.e. choose the set of predictor inputs such that $\mathcal{D}_j$ has property 1). The following is a rule for choosing which external inputs to include as predictor inputs (see [11] for more details). From Proposition 1 there may be terms related to external variables r_k, $k \in \mathcal{Z}_j$ appearing in the expression of w_j. This happens when there is a path from r_k, $k \in \mathcal{Z}_j$ to w_j that passes only through nodes w_ℓ, $\ell \in \mathcal{Z}_j$ [11]. If there are any such terms present in the expression for w_j, choose these external variables as predictor inputs (choose k to be in $\mathcal{P}_j$ for any r_k with this property).

B. Periodic External Excitation

In this subsection we present the first non-parametric method for identifying G^0_{ji} embedded in a dynamic network when there is periodic excitation present. The main mechanism is based on averaging over several periods of data: all the non-periodic components (i.e. the noise) are averaged away as increasing periods are used to estimate the periodic component of the signals. This method is advocated in [10] for open-loop, closed-loop and MIMO systems. Here we extend the method to identification in networks with known interconnection structure.

The data set is split into M segments, where each segment consists of an integer number of periods of the excitation signal(s). Then the periodic component of the predictor inputs and the output is estimated by averaging. This requires the following assumptions on the noise.

Assumption 2: The process noise and sensor noise variables have the following properties:

(a) All noise variables have zero mean.

(b) The correlation length of each noise variable is less than the selected periodic interval.

Condition (b) will typically only be approximately satisfied in practice. However, for long data records, it is usually possible to choose a sufficiently long periodic interval such that the condition is approximately satisfied. Let superscript (p) denote the periodic component of a signal. Let $w_j^{[\ell]}$ denote the ℓth data block of size M of w_j. The average of w_j over M segments of data, where each segment of data contains an integer number of periods is equal to

$$\bar{w}_j(t) = \frac{1}{M}\sum_{\ell=1}^{M} \tilde{w}_j^{[\ell]}(t). \qquad (7)$$

By Assumption 2 and the expression (5), the asymptotic expression for $\bar{w}_j$ is:

$$\lim_{M\to\infty}\mathbb{E}[\bar{w}_j]=\lim_{M\to\infty}\mathbb{E}\Big[\frac{1}{M}\sum_{\ell=1}^{M}\Big(\sum_{k\in\mathcal{D}_j}\breve{G}^0_{jk}(q)w_k^{[\ell]}(t) + \sum_{k\in\mathcal{Z}_j}\breve{F}^0_{jk}(q)r_k^{[\ell]}(t) + r_j^{[\ell]}(t) + \breve{v}_j^{[\ell]}(t) + s_j^{[\ell]}(t)\Big)\Big]$$
$$=\lim_{M\to\infty}\frac{1}{M}\sum_{\ell=1}^{M}\sum_{k\in\mathcal{D}_j}\breve{G}^0_{jk}(q)w_k^{(p)}(t)+\sum_{k\in\mathcal{Z}_j}\breve{F}^0_{jk}(q)r_k^{(p)}(t)+r_j^{(p)}(t).$$

For notational convienience, let $\breve{K}^0_j = [\breve{G}^0_{j\mathcal{D}}\ \breve{F}^0_{j\mathcal{P}}]$, $u(t) = [w_\mathcal{D}(t)\ r_\mathcal{P}(t)]^T$, and $\bar{y}(t) = \bar{w}_j(t) - r_j^{(p)}(t)$. Using this notation the above expression can be rewritten as:

$$\lim_{M\to\infty}\mathbb{E}[\bar{y}] = \breve{K}^0_j(q)u^{(p)}(t).$$

It follows that the cross-correlation between y and u is

$$\lim_{M\to\infty}\mathbb{E}[R_{\bar{y}u}(\tau)] = \breve{K}^0_j R_{u^{(p)}}(\tau) \qquad (8)$$

Note that $u^{(p)}$ is not known (since only a measured version of u is available), however, $u^{(p)}$ can be estimated by averaging (the same way as in (7)). By Assumption 2 it follows that $\lim_{M\to\infty}\mathbb{E}[\bar{u}(t)] = u^{(p)}(t)$. Thus, by taking the Fourier Transform of (8), $\breve{K}^0_j$ can be estimated as:

$$\hat{K}_j = \Phi_{\bar{y}\bar{u}}\Phi_{\bar{u}}^{-1} \qquad (9)$$

at any frequency where $\Phi_{\bar{u}}$ is invertible, where $\Phi_{\bar{u}}$ and $\Phi_{\bar{y}\bar{u}}$ are the Fourier Transforms of $R_{u^{(p)}}$ and $R_{\bar{y}u}$ respectively. From Proposition 1 if the predictor inputs are chosen such that they have Property 1 then (9) results in a consistent estimate of G^0_{ji}. If u is a vector of more than one predictor input then it may be necessary to have more than one external variable exciting the network. For instance, if two transfer functions are to be estimated, and two uncorrelated, periodic external variables are used to excite the network, then $\Phi_{\bar{u}}$ will be invertible at all frequencies where both inputs have non-zero power. If only one external variable is used, then $\Phi_{\bar{u}}$ will only be invertible at a limited number of frequencies. For more information on the design of multivariable inputs see [10] for instance.

C. Arbitrary Excitation

In this section we present a non-parametric *instrumental variable*-based approach to estimate G^0_{ji}. Instrumental variables (IVs) have been used in open-loop, closed-loop and network identification [12], [13], [10], [8] but typically in a parametric framework. By using instrumental variables, periodic external excitation is no longer strictly required. In fact, depending on the interconnection structure, external excitation may not be required at all (the process noise driving the network may provide sufficient excitation). The trade-off is that more restrictive assumptions are required on the noise properties to ensure consistent estimates. Sensor noise can still be dealt with. The presented method is based on a parametric version of the method presented in [8].

IV methods are especially attractive in a network setting because there are typically many variables that are measurable in the network. We consider any external or measured internal variable other than those selected as predictor inputs or the output w_j as candidate instrumental variables. Denote the set of internal variables selected as IVs as w_ℓ, $\ell \in \mathcal{I}_j$, and the set of external variables selected as IVs as r_ℓ, $\ell \in \mathcal{X}_j$. A variable selected as an instrumental variable is denoted z_ℓ. If multiple variables are selected as IVs, then $z = \sum_\ell z_\ell = \sum_{\ell\in\mathcal{I}_j} w_\ell + \sum_{\ell\in\mathcal{X}_j} r_\ell$. Each IV must have the following properties.

Property 2: Every IV z_ℓ satisfies the following conditions:

1) z_ℓ is correlated to w_i,
2) z_ℓ is not correlated to any of the sensor noise variables s_k, $k \in \mathcal{D}_j$,
3) s_j is uncorrelated to either z_ℓ or to all w_k, $k \in \mathcal{D}_j$,
4) $R_{z_\ell \breve{v}_j}(\tau) = 0$, $\tau \geq 0$.

In the following text we first show why these conditions are important, then we show how to select IVs that satisfy these properties. Due to Conditions (1) and (2) of Property 2, the cross correlation between an IV and a measured variable $\tilde{w}_k$, $k \in \mathcal{D}_j$ is equal to the cross correlation between the IV and the sensor noise free variable:

$$R_{\tilde{w}_k z}(\tau) = R_{w_k z}(\tau) + R_{s_k z}(\tau) = R_{w_k z}(\tau), \text{ for all } k \in \mathcal{D}_j.$$

Consequently, By Proposition 1

$$\begin{aligned} R_{\tilde{w}_j z}(\tau) &= \sum_{k\in\mathcal{D}_j} \breve{G}^0_{jk}(q) R_{w_k z}(\tau) + \sum_{k\in\mathcal{P}_j} \breve{F}^0_{jk}(q) R_{r_k z}(\tau) \\ &\qquad + R_{s_j z}(\tau) + R_{\breve{v}_j z}(\tau) \\ &= \sum_{k\in\mathcal{D}_j} \breve{G}^0_{jk}(q) R_{\tilde{w}_k z}(\tau) + \sum_{k\in\mathcal{P}_j} \breve{F}^0_{jk}(q) R_{r_k z}(\tau) + R_{\breve{v}_j z}(\tau) \\ &\qquad + R_{s_j z}(\tau), \end{aligned} \tag{10}$$

where in the first equality the 'input' is $R_{w_k z}$ which is unknown, but in the second equality the 'input' is $R_{\tilde{w}_k z}$ which can be estimated from the data. For notational simplicity, gather all the inputs in a matrix:

$$u(\tau) = \left[R_{\tilde{w}_{m_1} z}(\tau) \cdots R_{\tilde{w}_{m_n} z}(\tau)\ R_{r_{\ell_1} z}(\tau) \cdots R_{r_{\ell_p} z}(\tau)\right]^T$$

where $\{m_1, \ldots, m_n\} = \mathcal{D}_j$ and $\{\ell_1, \ldots, \ell_p\} = \mathcal{P}_j$. Let $y(\tau) = R_{\tilde{w}_j z}(\tau)$, and $s^R(\tau) = R_{s_j z}(\tau)$. As in Section III-B, let $\breve{K}_j = [\breve{G}^0_{j\mathcal{D}}(q)\ \breve{F}^0_{j\mathcal{P}}(q)]$. Now, (10) can be expressed as:

$$y(\tau) = \breve{K}^0_j u(\tau) + s^R_j(\tau), \quad \text{for } \tau \geq 0, \tag{11}$$

where $R_{\breve{v}_j z}(\tau)$ has been removed from the equation, because by Condition (4) it is equal to zero for all $\tau \geq 0$. Consider now, correlating the equation with u and taking the Fourier Transform of u and y for $\tau \geq 0$:

$$\Phi_{yu}(\omega) = \breve{K}^0_j \Phi_u(\omega) \tag{12}$$

where the term s^R is removed from the expression due to Condition (3) of Property 2. Thus, $\breve{K}^0_j$ can be estimated as:

$$\hat{K}_j = \Phi_{yu}(\omega) \Phi_u^{-1}(\omega) \tag{13}$$

at any frequency where Φ_u is invertible. Equation (13) results in consistent estimates of G^0_{ji} if $\mathcal{D}_j$ is chosen such that it has Property 1 and the IVs are chosen such that they have Property 2. Again, the invertibility of Φ_u poses a constraint. Note that in a practical situation Φ_u and Φ_{yu} cannot be exactly calculated and must be estimated. There exist many methods to obtain good estimates of (cross) power spectral densities as described in [9], [10] for instance.

The following proposition presents a more intuitive interpretation of Condition (4) of Property 2.

Proposition 2: Consider a dynamic network as defined in Section II. Suppose sets of predictor inputs and instrumental variables have been chosen. Condition (4) of Property 2 holds if the following conditions are satisfied:

(a) there is no path from w_j to any w_ℓ, $\ell \in \mathcal{I}_j$.
(b) there is no v_m, $m \in \mathcal{Z}_j$ that has both a path from v_m to w_j that passes only through w_k, $k \in \mathcal{Z}_j$ and a path from v_m to any w_ℓ, $\ell \in \mathcal{I}_j$. □

For a proof see Appendix V-A. If there are measured internal variables present in the network for which Condition (a) is satisfied, then external excitation is not strictly necessary. Note that external variables always satisfy both Conditions, and so they can always be used as IVs.

D. Noise Modelling

In the previous section the interconnection structure of the network was used to deal with the process noise (see Proposition 2), with the result that not every measured internal variable could be used as an instrumental variables. For the method we present in this section every measured internal variable (other than those selected as predictor inputs) can be used as an IV. The method works by using a noise model to take care of the disturbing process noise $\breve{v}_j$. It is known that a noise model is essential when using the so called Direct Method in closed-loop and network identification [14], [1]. A similar mechanism is employed in the method presented in this section. This method is a non-parametric method that is based on the parametric method of [8].

In the following text we will derive the expression for the non-parametric estimate. Although the final expression will be very similar to that of the previous section (i.e. Equation (13)) we will start the derivation by looking at the objective function to illustrate the use of noise modelling in a non-parametric framework. Consider the objective function:

$$V(\theta) = \frac{1}{F} \sum_{k=0}^{F-1} \left| H^{-1}(k) \big(Y(k) - \theta(k) U(k)\big) \right|^2, \tag{14}$$

where Y and U are the Discrete Fourier Transforms of $y(\tau)$ and $u(\tau)$ as defined just after (10), the argument k denotes the frequency point $e^{j2\pi k/N}$, $\theta(k)$ is the unknown transfer function at each frequency point, H is the noise model, and F is a user selected maximum frequency point. Equation (14) is simply the Fourier Transform of the standard one-step-ahead predictor [9]. It is well known that the noise model is equivalent to a weighting function in frequency [9], [10]. We briefly present the result using the reasoning of [10]. For the sake of argument, suppose (for now) that the noise model H is known. From (14) it follows

$$V(\theta) = \frac{1}{F} \sum_{k=0}^{F-1} \big(Y - \theta U\big) \frac{\sigma_e^2}{\Phi_{\breve{v}}} \big(Y^* - U^* \theta^*\big), \tag{15}$$

where the argument k has been dropped for clarity, $(\cdot)^*$ denotes complex conjugate transpose, and $\Phi_{\breve{v}}$ is the noise power spectral density and is equal to $\Phi_{\breve{v}} = \sigma^2_{\breve{e}} |H_j|^2$, where $\sigma^2_{\breve{e}}$ is the variance of the white noise source $\breve{e}_j$. Consider the notation for a weighted cross power spectral density:

$$\Phi^W_{yu}(k) = Y(k) W(k) U^*(k),$$

where W is the weighting function. Using this notation, (15) can be expressed as:

$$V(\theta) = \frac{1}{F}\sum_{k=0}^{F-1} \Phi_y^W - \theta\Phi_{yu}^W - \Phi_{uy}^W\theta^* + \theta\Phi_u^W\theta^*, \quad (16)$$

where $W(k) = \frac{\sigma_e^2}{\Phi_v(k)}$. Thus, the value of θ that minimizes (14) at each frequency can be calculated by taking the derivative and setting the resulting expression to zero. From (16) the resulting estimate is:

$$\theta(k) = \left(\Phi_u^W(k)\right)^{-1}\Phi_{uy}^W(k). \quad (17)$$

This expression can be evaluated as long as the weighting function W is known. In [10] two methods for estimating the noise model are presented: for periodic and arbitrary inputs respectively. We only show how to estimate the noise model for periodic inputs. For arbitrary excitation, the noise model can be estimated using the Local Polynomial Method [10].

Because the external excitation is periodic, the variance of the non-periodic component of the signal is purely due to noise. In [10] it is shown that if this variance is calculated at each frequency, the result is a non-parametric noise model. Let $\bar{Y}$ denote the Discrete Fourier Transform (DFT) of $\bar{y}(\tau), \tau \geq 0$, where the $\bar{\cdot}$ notation is defined in (7). Let $Y^{[\ell]}$ denote the DFT of the ℓth periodic block of $y(\tau)$, $\tau \geq 0$. Then the power spectral density of $\breve{v}_j$ is equal to [10]:

$$\hat{\Phi}_{\breve{v}}^2(k) = \frac{1}{M-1}\sum_{\ell=1}^{M} |Y^{[\ell]}(k) - \hat{Y}(k)|^2$$

Thus the noise power spectral density can be estimated, and then subsequently used as a weighting function in (17). The following proposition shows that if the predictor inputs are chosen such that they have Property 1, $\breve{v}_j$ is uncorrelated to the other noise sources in the network, then consistent estimates of G_{ji}^0 are possible using (17).

Proposition 3: Consider a dynamic network as defined in Section II. Suppose that the set of predictor inputs, $\mathcal{D}_j$ satisfies Property 1 and $\mathcal{P}_j$ is selected by the rules stated following Proposition 1. Then (17) is a consistent estimate of G_{ji}^0 if the following conditions are satisfied:

(a) The power spectral density matrix is full rank at all frequencies corresponding to $k = 0, \ldots, F-1$.
(b) Every sensor noise variable s_n, $n \in \mathcal{D}_j \cup \{j\}$ is uncorrelated to every s_ℓ, $\ell \in \mathcal{I}_j$.
(c) There is no v_m, $m \in \mathcal{Z}_j$ that has a path from v_m to w_j that passes only through w_k, $k \in \mathcal{Z}_j$ and a path from v_m to any w_ℓ, $\ell \in \mathcal{D}_j \cup \mathcal{I}_j$.
(d) All process noise variables are uncorrelated to each other.
(e) There is a delay in every path from w_j to w_ℓ, $\ell \in \mathcal{D}_j$.

For a proof see Appendix V-B.

Example 1: The dynamic network shown in Figure 1 is simulated. The objective is to estimate G_{21}^0 shown in green. Thus, w_2 is the output, w_1 is the predictor input, and we choose w_3 as the IV. We use the method based on non-parametric noise modelling, so we require r_1 to be periodic. The external variable, r_1 is chosen to be a *Pseudo Random Binary Sequence* of length 1000, repeated 100 times. The results of 10 independent experiments are shown in Figure 2. The estimate for each experiment is shown in gray. The estimated dynamics are consistent. The yellow lines are the estimate that results when no noise model is used. This is to show that without using a noise model, consistent estimates are not obtained.

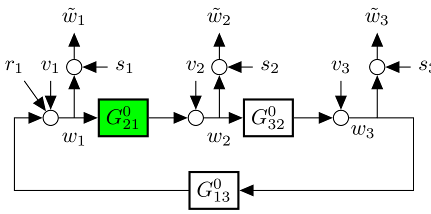

Fig. 1. A simple dynamic network simulated in Example 1.

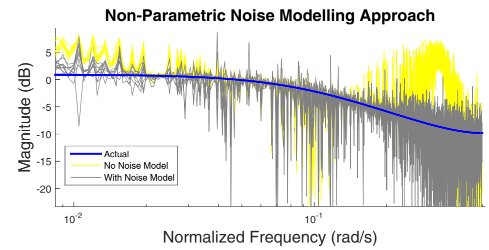

Fig. 2. Results of 10 simulation experiments using the noise modelling approach for the system shown in Fig. 1. The grey lines indicate the non-parametric estimates using a noise model, and the yellow lines show the estmates when no noise model is used.

IV. Conclusion

Non-parametric identification has many advantages in identification in dynamic networks. The methods require fewer assumptions on the noise, less knowledge about the system (in terms of choosing a parameterization) and the methods are generally less complex from an optimization point of view. These advantages become increasingly important when moving into a dynamic network framework, where noise is probably correlated, less is known about the system, and the systems become larger.

V. Appendices

A. Proof of Proposition 2

The following lemma is used in the proof. It can be proved using Mason's Rules or as shown in Appendix A of [1].

Lemma 1: Consider a dynamic network with transfer matrix G^0 that satisfies all conditions of Assumption 1. Let $\mathcal{G}_{mn}^0$ be the (m,n)th entry of $(I - G^0)^{-1}$. If all paths from n to m are zero (have a delay) then $\mathcal{G}_{mn}^0$ is zero (has a delay). □

Now consider the proof of Proposition 2.

Proof: Consider the expression (6) for $\breve{v}_j$. Suppose that w_ℓ is the IV. Express w_ℓ in terms of only v's and r's as:

$$w_\ell = \sum_{k=1}^{L} \mathcal{G}_{\ell k}^0(v_k + r_k), \quad (18)$$

where $\mathcal{G}_{\ell k}$ is the (ℓ,k)th entry of $(I-G^0)^{-1}$. By Condition (a) and Lemma 1 $\mathcal{G}_{\ell j}=0$. Consider each v_m that appears in both expressions (6) and (18). By Condtion (b) and Lemma 1 either F^0_{jm} in (6) or $\mathcal{G}^0_{\ell m}$ in (18) is zero for every v_m. Because all process noise variables are assumed uncorrelated, the expressions (6) and (18) are uncorrelated for $\tau \geq 0$. ■

B. Proof of Proposition 3

Proof: The proof proceeds by analyzing the function

$$V(\theta)=\mathbb{E}\Big[\frac{1}{2\pi}\int_{-\pi}^{\pi}\big(Y-\theta U\big)W\big(Y^*-U^*\theta^*\big)\mathrm{d}\omega\Big], \quad (19)$$

where the argument $e^{j\omega}$ is omitted for notational clarity, and $W=\frac{1}{\Phi_{\breve{v}}}$. The objective function (19) is the asymptotic version of (14) as $N\to\infty$, thus we can analyse the consistency of (14) using (19).

First we show that under the conditions of Proposition 3, $\breve{v}_j$ is causally uncorrelated to all other noise sources affecting the predictor inputs and the instrumental variables. By Proposition 1 an internal variable w_ℓ, $\ell\in\mathcal{D}_j$, can be expressed in terms of w_n, $n\in\{\mathcal{D}_j\cup\{j\}\}\setminus\{\ell\}$ as:

$$w_\ell=\sum_{n\in\mathcal{D}_j\cup\{j\}}\breve{G}^0_{\ell n}w_n+\sum_{n\in\mathcal{P}_j}\breve{F}^0_{\ell n}r_n+\breve{v}_\ell,$$

where $\breve{v}_\ell$ is the effective noise directly affecting w_ℓ. The noise variable v_m described in Condition (c) is called a *confounding variable*. In [11] it is shown that if there are no confounding variables, then $\breve{v}_j$ is uncorrelated to all other effective noise terms $\breve{v}_\ell$, $\ell\in\mathcal{D}_j$. The same reasoning can be applied to show that $\breve{v}_j$ is uncorrelated to the effective noise directly affecting the instrumental variables.

Additionally, in [11] it is shown that Condition (e) has the result that each term in w_j that is associated with a predictor input is not a function of $\breve{v}_j(t)$, but delayed versions of $\breve{v}_j(t)$. This result relies heavily on Lemma 1.

Now it is shown how these two facts are needed for the consistency result. By Proposition 1, $Y=\breve{K}^0_jU+\breve{X}_j$, where $\breve{X}_j=\mathcal{F}[R_{\breve{v}_jz}]$. Plugging this expression into (19) results in

$$V(\theta)=\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\big(\Delta K(\theta)U+\breve{X}_j\big)W\big(U^*\Delta K^*(\theta)+\breve{X}^*_j\big)\mathrm{d}\omega],$$

where $\Delta K=\breve{K}^0_j-\theta$. Let U in (19) be expressed as:

$$U=F_1R+F_2X_{\mathcal{D}}+F_3\breve{X}_j$$

where F_1R is the component of U due to the r terms, $F_2X_{\mathcal{D}}$ is the component of U due to the effective noise terms $\breve{v}_\ell$, $\ell\in\mathcal{D}_j$, and $F_3\breve{X}_j$ is the component of U due to $\breve{v}_j$. Then:

$$\begin{aligned}V=&\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(\Delta K(\theta)(F_1R+F_2X_{\mathcal{D}})+(\Delta K(\theta)F_3+1)\breve{X}_j\Big)\\&\cdot W\Big((F_1R+F_2X_{\mathcal{D}})^*\Delta K^*(\theta)+\breve{X}^*_j(1+F^*_3\Delta K^*(\theta))\Big)\mathrm{d}\omega]\\=&\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Delta K(\theta)(F_1R+F_2X_{\mathcal{D}})W(F_1R+F_2X_{\mathcal{D}})^*\Delta K^*(\theta)\mathrm{d}\omega]\\&+\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}(\Delta K(\theta)F_3+1)\breve{X}_jW\breve{X}^*_j(1+F^*_3\Delta K^*(\theta))\big)\mathrm{d}\omega]\end{aligned}$$

where the second equality holds because $\breve{X}_j$ is uncorrelated to $X_{\mathcal{D}}$ by the reasoning at the beginning of the proof. Consider now the second integral in the above expression for V. Plugging in $W=\Phi_{\breve{v}}=1/(\breve{X}_j\breve{X}^*_j)$ results in:

$$\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}(\Delta K(\theta)F_3+1)(1+F^*_3\Delta K^*(\theta))\big)\mathrm{d}\omega].$$

The terms $\Delta K(\theta)F_3$ and $F^*_3\Delta K^*(\theta)$ both have a factor $e^{j\omega}$ by the reasoning in the beginning in the proof which states that F_3 has a delay. The integral of any linear transfer function with a factor $e^{j\omega}$ is zero, and so these two terms disappear from the expression. Thus, finally,

$$\begin{aligned}V=&\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Delta K(\theta)(F_1R+F_2X_{\mathcal{D}})W(F_1R+F_2X_{\mathcal{D}})^*\Delta K^*(\theta)\mathrm{d}\omega]\\&+\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Delta K(\theta)F_3F^*_3\Delta K^*(\theta))\big)\mathrm{d}\omega]+\sigma^2_e\\=&\mathbb{E}[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Delta K(\theta)UWU^*\Delta K^*(\theta)\mathrm{d}\omega]+\sigma^2_e.\end{aligned}$$

Because Φ^W_u is full rank at all frequencies this expression is a minimum if and only if $\Delta K(\theta)=0$. ■

References

[1] P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods - basic methods for consistent module estimates," *Automatica*, vol. 49, pp. 2994–3006, Oct. 2013.

[2] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson, "Variance results for identification of cascade systems," *Automatica*, vol. 45, no. 6, pp. 1443 – 1448, 2009.

[3] N. Everitt, H. Hjalmarsson, and C. Rojas, "Variance results for parallel cascade serial systems," in *Proceedings of 19th IFAC World Congress*, Cape Town, South Africa, 2014, pp. 2317–2322.

[4] A. Haber and M. Verhaegen, "Identification of spatially distributed discrete-time state-space models," in *Proceedings of the 16th IFAC Symposium on System Identification*, Brussels, Belgium, 2012, pp. 410–415.

[5] D. Materassi, M. V. Salapaka, and L. Giarré, "Relations between structure and estimators in networks of dynamical systems," in *Proceedings of the 50th IEEE Conference on Decision and Control (CDC)*, Orlando, USA, 2011, pp. 162–167.

[6] J. Gonçalves and S. Warnick, "Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, Aug. 2008.

[7] Y. Yuan, G. Stan, S. Warnick, and J. Goncalves, "Robust dynamical network structure reconstruction," *Automatica*, vol. 47, no. 6, pp. 1230 – 1235, 2011.

[8] A. Dankers, P. M. J. Van den Hof, X. Bombois, and P. S. C. Heuberger, "Errors in variables identification in dynamic networks - consistency results for an instrumental variable approach," *Automatica*, 2015, accepted as regular paper.

[9] L. Ljung, *System Identification. Theory for the User*, 2nd ed. Prentice Hall, 1999.

[10] R. Pintelon and J. Schoukens, *System Identification, A Frequency Domain Approach*, 2nd ed. Hoboken, New Jersey, USA: IEEE Press, John Wiley and Sons, Inc., 2012.

[11] A. Dankers, P. M. J. Van den Hof, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods - predictor input selection," To appear in *IEEE Transactions on Automatic Control*, vol. 61, no. 4, April 2016.

[12] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification*, ser. LNCIS. Berlin: Springer, 1983.

[13] M. Gilson and P. M. J. Van den Hof, "Instrumental variable methods for closed-loop system identification," *Automatica*, vol. 41, pp. 241–249, 2005.

[14] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, pp. 1215–1241, 1999.