Frequency Domain Estimation of Parallel Hammerstein Systems using Gaussian Process Regression

Jeremy G. Stoddard * **James S. Welsh** *

* *School of Electrical Engineering and Computing, University of Newcastle, Australia (e-mail: jeremy.stoddard@uon.edu.au)*

Abstract: There exists a number of nonparametric model structures specifically developed for frequency domain modelling of nonlinear systems. Here we consider the nonlinear output frequency response function (NOFRF) structure, which is a series of input-dependent one-dimensional functions representing each nonlinear order present in the system. When used to model parallel Hammerstein systems, the NOFRFs lose their input dependence and become 'linear' in structure. In this paper, we extend a linear Gaussian process regression method to the nonlinear setting, where the pseudo-linear form of Hammerstein NOFRFs can be exploited by applying standard covariance structures from the linear theory. Compared to the traditional method of NOFRF estimation, the proposed method can be performed using simple experimental conditions and shows a significant improvement in estimation accuracy in the presence of measurement noise. The proposed method can also be adapted to estimate and remove the effect of transients in the case of non-periodic excitation. Numerical results are presented which show the veracity of the proposed algorithms for systems with polynomial nonlinearities of known degree.

Keywords: System identification, Nonlinear systems, Frequency domain, Regularization

1. INTRODUCTION

In the time domain, nonlinear systems can be described nonparametrically using a Volterra series representation (Schetzen, 1980), where each term in the series represents a different nonlinear order. In the frequency domain however, there are several competing representations which offer different benefits for interpretation and estimation (Rijlaarsdam et al., 2017), (Cheng et al., 2017).

One of the more natural approaches in the frequency domain is to apply multidimensional Fourier transforms to each Volterra series term, producing the so-called Generalized Frequency Response Functions (GFRFs) (George, 1959). Due to the complicated multidimensional nature of the GFRFs, an alternate model structure was proposed by Lang and Billings (2005) which contains only one-dimensional frequency functions. These new quantities, known as 'nonlinear output frequency response functions' (NOFRFs), have a more intuitive interpretation from linear system theory, however the functions are input dependent in general. The representation has found use in fault detection applications (Peng et al., 2007), (Cao et al., 2013), (Xia et al., 2015) as well as nonlinearity detection (Lang and Peng, 2008).

The NOFRF model proposal in Lang and Billings (2005) was accompanied by a data-driven identification algorithm we call the 'traditional method', which requires multiple experiments on the system under conditions that may not be possible in practice, and employs a least squares solution that will be sensitive to noisy measurements. In the special case of parallel Hammerstein systems, however, it may be possible to perform more sophisticated estimation, since the NOFRFs become independent of the applied input, and the form of the functions closely resembles that of the underlying linear filters in the system. This property allows us to borrow well-developed concepts from linear identification theory.

In this paper, we develop a Gaussian process regression method for estimating Hammerstein NOFRFs, which can be seen as an extension to the linear frequency domain method outlined in Lataire and Chen (2016). The linear method is itself a frequency domain interpretation of the popular Bayesian regularization approach for impulse response estimation, introduced in Pillonetto and De Nicolao (2010). By framing the NOFRFs as normally distributed quantities with standard prior covariance structures, we can estimate all NOFRFs in a regularized fashion using only one experiment and with less restrictions than the traditional method. Numerical examples reveal that the newly proposed method is also significantly less sensitive to measurement noise than traditional estimation. The more general case of non steady-state data is also considered, where the regularized approach can be adapted to minimize the effect of transients on NOFRF estimation.

The proposed method is distinct from other parallel Hammerstein identification algorithms in that it is a fully nonparametric method applied directly in the frequency domain. Traditionally, the linear dynamic and nonlinear blocks could be estimated separately and parametrically in an iterative scheme (Gallman, 1975; Schoukens et al.,

2011), however the resulting models do not provide direct intuition on frequency domain behaviour. More recently, Gaussian process regression has been applied to Hammerstein identification in the time domain (Risuleo et al., 2017), but the approach was limited to a single Hammerstein branch.

The paper is organised as follows. Section 2 defines the necessary notation for complex normal distributions. Section 3 presents the NOFRF model and traditional identification scheme from Lang and Billings (2005). Section 4 develops the regularized method using Gaussian process regression, and the newly proposed method is evaluated alongside the original algorithm through numerical simulations in Section 5. Finally, conclusions are drawn in Section 6.

2. COMPLEX NORMAL DISTRIBUTIONS

A complex random vector, X, is said to be complex normally distributed if its real and imaginary parts are normally distributed. The properties of complex normally distributed quantities are largely the same as for the real case, and the interested reader is directed to Schreier and Scharf (2010) for a thorough treatment. In this section, we define the notation adopted in this paper, and present the relevant properties for Gaussian process regression.

Definition 1. (Augmented vector). For a complex column vector, X, the augmented vector is defined as $\widetilde{X} = [X^T X^H]^T$, where T and H denote the transpose and Hermitian transpose respectively.

Definition 2. (Augmented mean and covariance). A complex normally distributed vector, X, is denoted by

$$X \sim \mathcal{CN}(\mu, \Sigma),$$

where $\mu = \mathbf{E}\{\widetilde{X}\}$ is the augmented mean, and $\Sigma = \mathbf{E}\{(\widetilde{X} - \mu)(\widetilde{X} - \mu)^H\}$ is the augmented covariance, which can be decomposed into covariance and relation functions, K and C, as

$$\Sigma = \begin{bmatrix} K & C \\ C^H & \overline{K} \end{bmatrix}$$

$$\text{with } K = \mathbf{E}\{(X - \mathbf{E}\{X\})(X - \mathbf{E}\{X\})^H\}$$

$$\text{and } C = \mathbf{E}\{(X - \mathbf{E}\{X\})(X - \mathbf{E}\{X\})^T\}.$$

Definition 3. (Complex circular distributions). A complex random vector, V, is complex circular if and only if $\mu = 0$ and $C = 0$.

For the following properties, consider $A \sim \mathcal{CN}(\mu_A, \Sigma_A)$ and $B \sim \mathcal{CN}(\mu_B, \Sigma_B)$

Property 1. (Independent Sum). For A and B independent and of equal dimension, their sum is distributed as

$$A + B \sim \mathcal{CN}(\mu_A + \mu_B, \Sigma_A + \Sigma_B). \tag{1}$$

Property 2. (Hadamard Product). For a complex vector U with equal dimension to A, the Hadamard product $U \circ A$ is distributed as

$$U \circ A \sim \mathcal{CN}(\widetilde{U} \circ \mu_A, (\widetilde{U}\widetilde{U}^H) \circ \Sigma_A). \tag{2}$$

Property 3. (Conditional Distributions). If A and B are jointly (Gaussian) distributed, the conditional distribution of A given B is given by

$$A|B \sim \mathcal{CN}\big(\mu_A + \Sigma_{AB}\Sigma_B^{-1}(\widetilde{B} - \mu_B), \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA}\big), \tag{3}$$

where $\Sigma_{AB} = \mathbf{E}\{(\widetilde{A} - \mu_A)(\widetilde{B} - \mu_B)\}$.

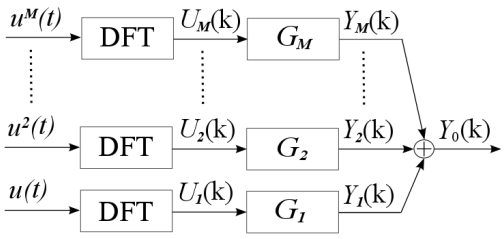

Fig. 1. Equivalent block structure of the NOFRF model

3. THE NOFRF MODEL

3.1 Model definition

Let $u(t)$ and $y_0(t)$ be discrete, noiseless input and output measurements from a nonlinear system, where $t = 0, 1, \ldots, N-1$. The N-point discrete Fourier transforms (DFTs) of u and y_0 will be labelled $U(k)$ and $Y_0(k)$, where k indicates the frequency bin. The NOFRF model definition in Lang and Billings (2005) assumes that the frequency domain measurements are transient-free (estimation in the presence of transients is discussed here in Section 4.2), i.e. the input to the system is periodic in N and has been applied for an infinitely long period before measurements were taken. In this case, the noiseless steady state output spectrum is given by,

$$\begin{aligned} Y_0(k) &= \sum_{m=1}^{M} Y_m(k) = \sum_{m=1}^{M} G_m(j\omega_k) U_m(k) \\ \text{where } U_m(k) &= \frac{1/\sqrt{m}}{(2\pi)^{m-1}} \sum_{k_1+\ldots+k_m=k} \prod_{i=1}^{m} U(k_i). \end{aligned} \tag{4}$$

In (4), $G_m(j\omega_k)$ is the m'th order NOFRF at frequency ω_k given by the k'th DFT bin. The quantities $U_m(k)$ are nonlinear extensions of the input spectrum, which are seen to be DFTs of the input raised to the m'th power, i.e. $u^m(t)$. The equivalent NOFRF block structure is given in Figure 1, showing the intuitive interpretation of NOFRFs as 'filters' on the nonlinear input quantities.

In this paper, we will also consider the effect of white output measurement noise on the identification. Thus, the observed output spectrum is given by

$$Y(k) = Y_0(k) + V(k), \tag{5}$$

where $V(k)$ is assumed to be complex circular noise with covariance $\sigma_v^2 I$, where I is the identity matrix.

3.2 The Hammerstein case

The NOFRFs are input-dependent in general, since they act on exponents of the input rather than the input directly. In the special case of a Hammerstein system with polynomial nonlinearity (Figure 2), the frequency functions become independent of the applied input and will be scaled versions of the system's linear filter, $H(j\omega)$. The reasoning can be easily seen by comparing Figure 1 with the rearranged Hammerstein structure in Figure 3. The two block structures will be equivalent if we set $G_m(j\omega) = a_m H(j\omega)\ \forall m$. The same logic also applies to parallel Hammerstein systems, where the NOFRF at each nonlinear order is formed from a linear combination of the filters in each parallel branch.

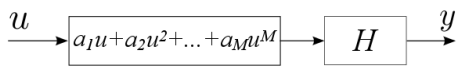

Fig. 2. Block structure of a Hammerstein system with polynomial nonlinearity

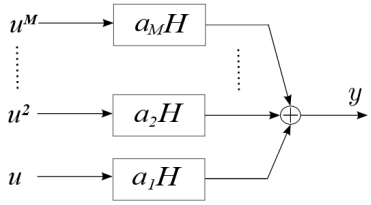

Fig. 3. Alternate block structure for a Hammerstein system

For Hammerstein and parallel Hammerstein systems, the NOFRFs can be treated as linear filters for identification, and the resulting model will be valid regardless of the system input. This is the case which will be considered in this paper.

3.3 Identification using multi-level excitation

Lang and Billings (2005) proposed a data-based identification method which requires the system to be excited multiple times by scaled versions of a prototype input signal. Labelling the unscaled input as $u^*(t)$, the method requires the system to be excited q times by the input signals,

$$\alpha_i u^*(t), \quad i = 1, \ldots, q,$$

where $q \geq M$ and $\alpha_1, \ldots, \alpha_q$ are positive constants. The excitations will produce q corresponding output frequency responses which are denoted as $Y_q(k)$.

Given the model structure in (4), the NOFRFs can be obtained by forming a linear system for each excited frequency, i.e.

$$\begin{bmatrix} Y_1(k) \\ \vdots \\ Y_q(k) \end{bmatrix} = \begin{bmatrix} \alpha_1 U_1(k) & \ldots & \alpha_1^M U_M(k) \\ \vdots & \ddots & \vdots \\ \alpha_q U_1(k) & \ldots & \alpha_q^M U_M(k) \end{bmatrix} \begin{bmatrix} G_1(j\omega_k) \\ \vdots \\ G_M(j\omega_k) \end{bmatrix}, \quad (6)$$

where $U_1, \ldots, U_M$ are the nonlinear spectral quantities corresponding to $u^*(t)$. Solving each system in a least squares sense will provide NOFRF estimates at the excited frequencies.

There are several disadvantages of this traditional estimation method:

- Data collection may take a long time to perform, requiring at least M experiments which must reach steady state before measuring.
- The method cannot be used in cases where the input is not precisely controlled.
- The least squares solutions will be sensitive to measurement noise.
- Choices for the quantity and value of scaling parameters, α_i, are arbitrary in nature.

4. GAUSSIAN PROCESS REGRESSION FOR NOFRFS

Taking inspiration from the linear case in Lataire and Chen (2016), a regularized estimation method can be developed within the Bayesian perspective, with a (complex) Gaussian assumption on our unknown quantities. To achieve this, some assumptions and notation must first be clarified.

Assumption 1. (System structure). The system of interest admits a parallel Hammerstein structure, where the nonlinear blocks are accurately represented by polynomials with a known maximum degree, M. Thus, there exist nonzero NOFRFs up to the M'th nonlinear order.

Remark 1. In the case where M is unknown, or the nonlinearities are not polynomial in nature, there are many established order selection methods, e.g. cross-validation, which can be applied to select an appropriate M.

Notation 1. Let $\mathbf{k}$ be the vector of DFT bins excited by the input, such that $\omega_{\mathbf{k}}$ are the corresponding excited frequencies. The parameter vector to be estimated for the m'th order NOFRF will then be $G_m(j\omega_{\mathbf{k}})$.

Assumption 2. (Independent Gaussian NOFRFs). Assume that each NOFRF parameter vector is a Gaussian process with zero mean, i.e.

$$G_m(j\omega_{\mathbf{k}}) \sim \mathcal{CN}(0, \alpha_{G,m}\Sigma_m), \quad m = 1, \ldots, M, \quad (7)$$

where $\alpha_{G,m} > 0$ and Σ_m is constructed from covariance and relation functions, K_m and C_m. Furthermore, assume that all NOFRFs are independent of each other.

Recalling that parallel Hammerstein NOFRFs can be interpreted as linear filters, we are free to use the same covariance structures for Σ_m as were used in Lataire and Chen (2016). Here we consider the frequency domain equivalent of the Diagonal Correlated (DC) kernel, which was originally developed for Bayesian impulse response estimation (Pillonetto et al., 2014). For continuous time systems, the frequency domain covariance is given by

$$\begin{aligned} K_m(j\omega_x, j\omega_y) &= \frac{1}{b_m + j(\omega_x - \omega_y)} \ldots \\ &\quad \times \Big(\frac{1}{a_m + b_m/2 + j\omega_x} + \frac{1}{a_m + b_m/2 - j\omega_y} \Big), \\ C_m(j\omega_x, j\omega_y) &= K_m(j\omega_x, -j\omega_y), \end{aligned} \quad (8)$$

where $a_m, b_m > 0$ are tunable hyperparameters describing the m'th order frequency function.

Theorem 1. For the NOFRF model defined in (4) and (5) with Gaussian G_m given by (7), the vectorized output spectrum, $Y(\mathbf{k})$, is distributed as,

$$\begin{aligned} Y(\mathbf{k}) &\sim \mathcal{CN}(0, \Sigma_Y), \\ \text{where } \Sigma_Y &= \sum_{m=1}^{M} \Sigma_{Y,m} + \sigma_v^2 I \\ \text{and } \Sigma_{Y,m} &= \alpha_{G,m}(\widetilde{U_m(\mathbf{k})}\widetilde{U_m(\mathbf{k})}^H) \circ \Sigma_m \end{aligned} \quad (9)$$

Proof 1. Follows directly from Assumption 2 and Properties 1 and 2 □

The NOFRF vectors are now jointly distributed with $Y(\mathbf{k})$, and the Bayesian framework allows computation of maximum a posteriori (MAP) estimates for each G_m.

Theorem 2. The MAP estimate of $G_m(j\omega_{\mathbf{k}})$ is given by,

$$\hat{G}_m(j\omega_{\mathbf{k}}) = \alpha_{G,m}\Sigma_m \mathrm{diag}\big(\widetilde{U_m(\mathbf{k})}^H\big)\Sigma_Y^{-1}\widetilde{Y(\mathbf{k})}, \quad (10)$$

where $\mathrm{diag}(X)$ denotes a diagonal matrix with diagonal X.

Proof 2. Acknowledging the independence of each NOFRF, the m'th order NOFRF and output spectrum are jointly distributed as,

$$\begin{bmatrix} G_m(j\omega_{\mathbf{k}}) \\ Y(\mathbf{k}) \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_m & \Sigma_{G_m Y} \\ \Sigma_{Y G_m} & \Sigma_Y \end{bmatrix} \right), \tag{11}$$

where the joint covariance is computed as,

$$\begin{aligned} \Sigma_{G_m Y} &= \mathbf{E}\Big\{ \widetilde{G_m(j\omega_{\mathbf{k}})} \widetilde{Y(\mathbf{k})}^H \Big\} \\ &= \mathbf{E}\Big\{ \widetilde{G_m(j\omega_{\mathbf{k}})} \Big[\widetilde{G_m(j\omega_{\mathbf{k}})}^H \circ \widetilde{U_m(\mathbf{k})}^H \Big] \Big\} \\ &= \alpha_{G,m} \Sigma_m \mathrm{diag}\big(\widetilde{U_m(\mathbf{k})}^H \big), \end{aligned} \tag{12}$$

using (4), (5) and Assumption 2. Now the MAP estimate will be the mean of the conditional distribution, $G_m(j\omega_{\mathbf{k}})|Y(\mathbf{k})$, which is given by Property 3 as

$$\mathbf{E}\Big\{ \widetilde{G_m(j\omega_{\mathbf{k}})} | \widetilde{Y(\mathbf{k})} \Big\} = \Sigma_{G_m Y} \Sigma_Y^{-1} \widetilde{Y(\mathbf{k})}. \tag{13}$$

Combining (13) with (12) yields the result in (10). □

4.1 Covariance hyperparameter tuning

The hyperparameters describing the covariance of each NOFRF must be tuned on the available data prior to estimation. For an M'th order model, the DC hyperparameter vector will be

$$\begin{aligned} \eta &= [\sigma_v^2 \ \eta_1 \ \dots \ \eta_M], \\ \eta_m &= [\alpha_{G,m} \ a_m \ b_m]. \end{aligned}$$

As in the linear case, the tuning process can be achieved by maximizing the log marginal likelihood of the hyperparameters with respect to the observed output (Lataire and Chen, 2016), i.e.

$$\hat{\eta} = \arg\min_{\eta} \widetilde{Y(\mathbf{k})}^H \Sigma_Y^{-1}(\eta) \widetilde{Y(\mathbf{k})} + \log\det \Sigma_Y(\eta). \tag{14}$$

4.2 Estimation in the presence of transients

Unlike the traditional multi-level excitation method, the regularized approach can be modified to remove the effect of transient functions in estimation, allowing identification of NOFRFs in the case of non-periodic or non-steady-state measurement data.

It was shown in Lataire and Chen (2016) that in the linear case, transient functions result from the difference $u_d(t) = u(t) - u(t+N)$ for $-\infty < t < 0$. Considering now the NOFRF case and observing Figure 1, it is clear that for a non-periodic $u(t)$, $u^m(t)$ will also be non-periodic, and there will be a transient function at each nonlinear order. We can use this information to update the model equation in (4) as follows:

$$Y_0(k) = \sum_{m=1}^{M} G_m(j\omega_k) U_m(k) + T_m(k), \tag{15}$$

where T_m is the m'th order transient function.

For a white noise input $u(t)$, it was shown in the linear case that the transient covariance is approximately proportional to the system covariance (Lataire and Chen, 2016). This result also extends to the NOFRF case for parallel Hammerstein systems.

Assumption 3. The distribution of the m'th order transient function can be approximately modelled as,

$$T_m(j\omega_{\mathbf{k}}) \sim \mathcal{CN}(0, \alpha_{T,m} \Sigma_m), \tag{16}$$

where $\alpha_{T,m} \in \mathbb{R}^+$ is a scaling hyperparameter.

Assumption 4. Assume that the Gaussian transient functions are independent of each other and all NOFRFs.

The distribution of the output spectrum can now be updated accordingly, giving

$$\begin{aligned} \Sigma_Y &= \sum_{m=1}^{M} \Sigma_{Y,m} + \sigma_v^2 I, \\ \Sigma_{Y,m} &= \alpha_{G,m} (\widetilde{U_m(\mathbf{k})} \widetilde{U_m(\mathbf{k})}^H) \circ \Sigma_m + \alpha_{T,m} \Sigma_m. \end{aligned} \tag{17}$$

Hyperparameter tuning can still be performed using (14), with additional hyperparameters in the optimization, i.e.

$$\eta_m = [\alpha_{G,m} \ \alpha_{T,m} \ a_m \ b_m].$$

Finally, the MAP estimates for each $G_m(j\omega_{\mathbf{k}})$ will remain as in Theorem 2, using the updated definition of Σ_Y in (17). Furthermore, MAP estimates of the transient functions can also be obtained via the same logic, yielding

$$\hat{T}_m(j\omega_{\mathbf{k}}) = \alpha_{T,m} \Sigma_m \Sigma_Y^{-1} \widetilde{Y(\mathbf{k})}. \tag{18}$$

5. NUMERICAL EXAMPLES

5.1 Simulation settings

Numerical simulations were performed to compare the performance of the proposed Gaussian process regression or 'regularized' method against the traditional method of multi-level excitation. All simulations used the parallel Hammerstein structure given in Figure 4, where each parallel branch represents a single nonlinear order and can be switched in or out of the system. With such a structure, the NOFRFs are directly equal to the linear filter in their corresponding branch.

The filters, G_1, G_2 and G_3, are constructed as 2nd order Chebyshev filters with resonant modes at 14.4 Hz, 22.1 Hz and 18.2 Hz respectively. The sampling frequency is set at 200 Hz, and the input in any given experiment is a random-phase multisine which is periodic in $N = 128$ samples. The band of excitation for the multisine is constructed so as to avoid aliasing in the output spectrum. The output error, e, is a Gaussian white noise vector added in each experiment to provide the required signal-to-noise ratio (SNR).

5.2 Comparison of traditional and regularized methods

Three 1000-run Monte Carlo studies were performed on the two estimation methods to determine their relative accuracy. The studies were:

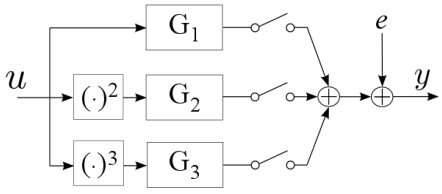

Fig. 4. Parallel Hammerstein structure used for numerical examples

(1) $M = 2$ (G_1 and G_2 branches) with SNR = 40dB
(2) $M = 2$ (G_1 and G_2 branches) with SNR = 20dB
(3) $M = 3$ (all branches) with SNR = 40dB

Note that while both methods used periodic multisines as input, the multi-level excitation method used the minimum requirement of M differently scaled inputs, requiring M experiments. In contrast, the regularized method used only one input and experiment, such that data collection is faster by a factor of M, and the estimation is performed on a dataset which is smaller by a factor of M.

For the comparison, all estimation data were taken from the steady-state portion of the experiment, i.e. there were no transient components in the output response.

To visualise performance in each Monte Carlo study, Figures 5, 6 and 7 plot the true NOFRF magnitudes alongside the central 95% intervals (shaded) from each estimation method. For all studies, the regularized method produced tighter intervals which are more centred around the true NOFRF, despite using significantly less data in the estimation. The results show a clear accuracy benefit for the proposed method when output measurement noise is present in the system, which can be attributed to the ablity of regularized methods to significantly reduce estimation variance at the price of a small bias (Pillonetto et al., 2014). The bias here is small but evident, particularly at the resonance peaks of each NOFRF.

5.3 Regularized estimation with transients

While the traditional multi-level excitation method cannot estimate or remove the effect of transients, it was shown in Section 4.2 that the regularized approach can be modified to consider transients in the Gaussian process regression. This allows us to estimate the NOFRFs and their corresponding transients, thereby increasing the accuracy of estimation for non-periodic or non steady-state data.

To demonstrate this capability, the modified method in Section 4.2 was applied to the simulated system of Figure

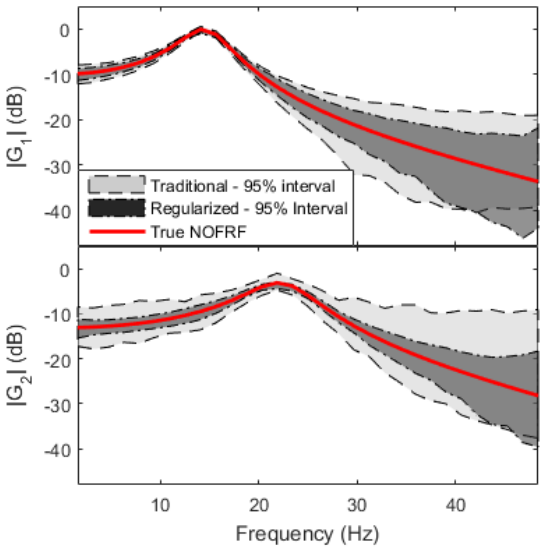

Fig. 5. Magnitude intervals for $M = 2$ and SNR=40dB

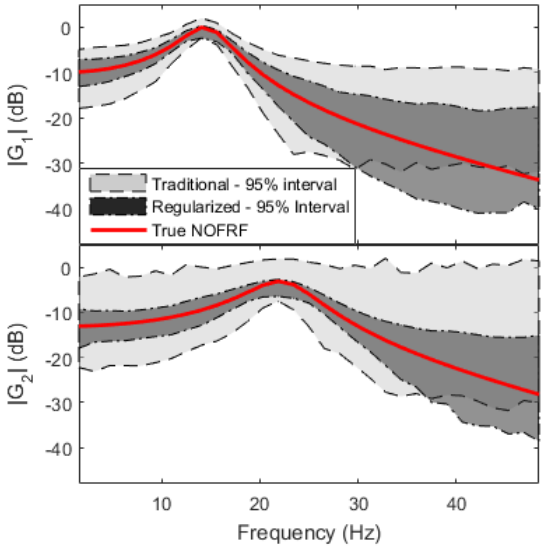

Fig. 6. Magnitude intervals for $M = 2$ and SNR=20dB

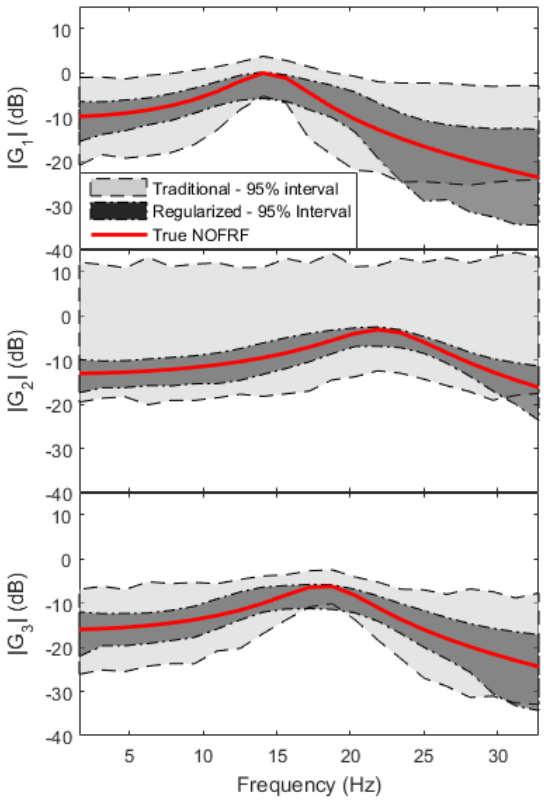

Fig. 7. Magnitude intervals for $M = 3$ and SNR=40dB

4, using the G_1 and G_2 branches and no output noise. The input multisine was applied for a single period and preceded by zero initial conditions, producing a transient in the output response. Using this input/ouput data, the total transient could be estimated by summing the transient estimates from each order, i.e.

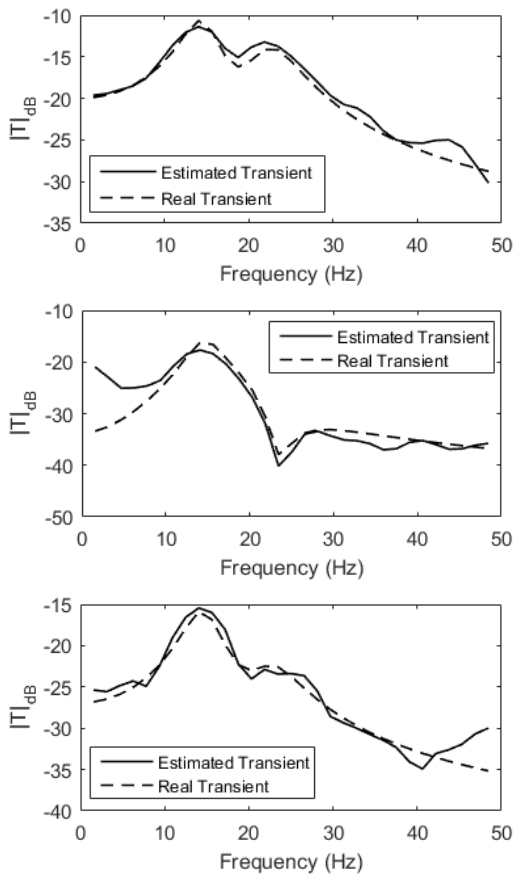

Fig. 8. True and estimated transient functions for three input realizations of the test system ($M = 2$)

$$T(j\omega) = \sum_{m=1}^{M} T_m(j\omega)$$

The true and estimated transient are plotted in Figure 8 for three input realizations, showing reasonable estimation performance which significantly reduces the effect of the transients on NOFRF estimation accuracy.

6. CONCLUSION AND FUTURE WORK

This paper demonstrates the benefits of a Gaussian process regression approach to frequency domain estimation of parallel Hammerstein systems. Using the NOFRF model structure, prior covariances for each frequency function can be chosen and tuned according to linear identification theory, due to the linear and input-invariant form adopted by Hammerstein NOFRFs. When compared to the traditional estimation method of multi-level excitation, the proposed method has fewer experimental constraints and can achieve higher accuracy when output noise is present. The method can also be adapted to estimate and remove the effect of transients resulting from non steady-state experiment data.

REFERENCES

Cao, J., Chen, L., Zhang, J., and Cao, W. (2013). Fault diagnosis of complex system based on nonlinear frequency spectrum fusion. *Measurement*, 46(1), 125 – 131.

Cheng, C.M., Peng, Z.K., Zhang, W.M., and Meng, G. (2017). Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review. *Mechanical Systems and Signal Processing*, 87(A), 340–364.

Gallman, P. (1975). An iterative method for the identification of nonlinear systems using a Uryson model. *IEEE Transactions on Automatic Control*, 20(6), 771–775.

George, D. (1959). *Continuous Nonlinear Systems*. MIT RLE Technical Report.

Lang, Z.Q. and Billings, S.A. (2005). Energy transfer properties of non-linear systems in the frequency domain. *International Journal of Control*, 78(5), 345–362.

Lang, Z. and Peng, Z. (2008). A novel approach for nonlinearity detection in vibrating systems. *Journal of Sound and Vibration*, 314(3), 603 – 615.

Lataire, J. and Chen, T. (2016). Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, 72, 217 – 229.

Peng, Z., Lang, Z., and Billings, S. (2007). The nonlinear output frequency response function and its application to fault detection. In *Fault Detection, Supervision and Safety of Technical Processes 2006*, 36 – 41. Elsevier Science Ltd, Oxford.

Pillonetto, G. and De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, 46(1), 81–93.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3), 657–682.

Rijlaarsdam, D., Nuij, P., Schoukens, J., and Steinbuch, M. (2017). A comparative overview of frequency domain methods for nonlinear systems. *Mechatronics*, 42(Supplement C), 11 – 24.

Risuleo, R.S., Bottegal, G., and Hjalmarsson, H. (2017). A nonparametric kernel-based approach to Hammerstein system identification. *Automatica*, 85, 234 – 247.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley and Sons.

Schoukens, M., Pintelon, R., and Rolain, Y. (2011). Parametric identification of parallel Hammerstein systems. *IEEE Transactions on Instrumentation and Measurement*, 60(12), 3931–3938.

Schreier, P. and Scharf, L. (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press.

Xia, X., Zhou, J., Li, C., and Zhu, W. (2015). A novel method for fault diagnosis of hydro generator based on NOFRFs. *International Journal of Electrical Power and Energy Systems*, 71(Supplement C), 60 – 67.