
Article:
Florescu, D. and Coca, D. (2018) Identification of linear and nonlinear sensory processing circuits from spiking neuron data. Neural Computation, 30 (3). pp. 670-707. ISSN 0899-7667

https://doi.org/10.1162/neco_a_01051

Identification of linear and nonlinear sensory processing circuits from spiking neuron data

Dorian Florescu, Daniel Coca*

Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, UK.

Keywords: System identification, spiking neural circuits, time encoding machines, time decoding machines, integrate-and-fire neurons

Abstract

Inferring mathematical models of sensory processing systems directly from input-output observations, while making the fewest assumptions about the model equations and the type of measurements available, is still a major issue in computational neuroscience. This paper introduces two new approaches for identifying sensory circuit models consisting of linear and nonlinear filters in series with spiking neuron models, based only on

the sampled analogue input to the filter and the recorded spike train output of the spiking neuron. For an ideal integrate-and-fire neuron model the first algorithm can identify the spiking neuron parameters as well as the structure and parameters of an arbitrary nonlinear filter connected to it. The second algorithm can identify the parameters of the more general, leaky integrate-and-fire spiking neuron model as well as the parameters of an arbitrary linear filter connected to it. Numerical studies involving simulated and real experimental recordings are used to demonstrate the applicability and to evaluate the performance of the proposed algorithms.

1 Introduction

System identification is widely used to develop quantitative models of sensory neurophysiology (Wu et al., 2006). The neural behaviour can be reproduced accurately using a wide range of models with various levels of complexity (Koch & Segev, 1998; Gabbiani & Cox, 2010). The sensory processing circuits, consisting of receptive fields and spiking neurons, have often been represented as cascade models, which aim to capture the key processing steps from the measured data (Herz et al., 2006). These models represent the receptive field as a filter that is linear (Paninski, 2004; Paninski et al., 2004; Lazar & Slutskiy, 2014) or nonlinear, satisfying the fading memory requirement (Lazar & Slutskiy, 2015; Song et al., 2016). The spiking neuron in a cascade model was represented by a threshold device with a feedback after-potential (Song et al., 2016), a static nonlinearity in series with a Poisson spike generator (Simoncelli et al., 2004), an integrate-and-fire (IF) neuron (Lazar & Slutskiy, 2015; Paninski et al., 2004), or a detailed Hodgkin-Huxley (HH) model (Lazar & Slutskiy, 2014). The linear-nonlinear-

Poisson (LNP) cascade model was extended to the generalized linear model (GLM), which includes additionally a feedback filter (Paninski, 2004). Other cascade model architectures can be found in (Hunter & Korenberg, 1986; Herz et al., 2006; Keat et al., 2001).

The integrate-and-fire (IF) neuron is one of the most common models of the spiking neuron (Lapicque, 1907; Tuckwell, 1988). The IF model has been shown to be a good approximation for biophysically detailed models like the Hodgkin-Huxley neuron (Kistler et al., 1997; Lazar & Slutskiy, 2010), as well as a good predictor for electrophysiological recordings (Clopath et al., 2007). There are two main classes of IF models: the ideal IF (IIF) and the more general leaky IF (LIF). Several variations of this model are presented in (Burkitt, 2006).

A popular identification methodology for sensory circuits estimates the LNP as well as the GLM model by maximizing a likelihood function depending on the model parameters (Simoncelli et al., 2004; Paninski, 2004; Pillow, 2007). This method was extended to cascade models comprising a linear filter in series with a variation of the LIF neuron with a feedback filter (Paninski et al., 2004). The maximum likelihood estimation of this model was performed successfully using extracellularly recorded spike train responses of the primate retinal ganglion cells to light stimuli (Pillow et al., 2005). Here, the threshold parameter δ of the LIF neuron is considered to be known *a priori*. Moreover, there is no detection routine performed to determine the structure of the filter, which is assumed to be known. A review on various identification methods for IF neurons can be found in (Burkitt, 2006).

Lazar & Tóth (2003) have proven that the IF neuron is a type of time encoding

machine (TEM) that converts the amplitude of an input signal into a sequence of spike times. The identification of neural circuits comprising filters in series with spiking neurons was formulated as an input reconstruction problem (Lazar & Slutskiy, 2015). More specifically, a method to identify circuits comprising of a linear filter in series with an IIF neuron (LF-IIF) was proposed by Lazar & Slutskiy (2010). By making additional assumptions, the identification approach has been extended to circuits where the IIF neuron is replaced by the LIF neuron (Lazar & Slutskiy, 2010) as well as the Hodgkin-Huxley (HH) model (Lazar & Slutskiy, 2014). In the first case, it is assumed that the LIF neuron parameters are known. In the second case, input-output measurements of the HH neuron are assumed to be available. The identification framework was extended further to circuits consisting of a nonlinear filter in series with an IIF neuron (NF-IIF), under the assumption that the filter admits a Volterra series representation (Lazar & Slutskiy, 2015). Another approach estimates multiple-input multiple-output (MIMO) generalized Volterra models, consisting of Volterra models in series with threshold devices with feedback after-potentials (Song et al., 2016). To solve the problem caused by the large number of coefficients, a group regularized estimation method is used to identify the model. This model was shown to predict accurately the spike trains from the hippocampal region CA1 based on spike train inputs recorded from CA3 during multiple memory events, making it suitable for implementation on a hippocampal memory prosthesis (Song et al., 2016).

The identification methods summarized above can accommodate a wide range of filters and spiking neurons. However, the assumptions made, such as the availability of input-output data from the spiking neuron or the *a priori* knowledge of spiking neuron

parameters, limit to some extent their practical applicability. Furthermore, if the filter is assumed to be nonlinear, which is often the case in practice, the direct identification of Volterra kernels has well known practical limitations (Chen & Billings, 1989).

This paper introduces two approaches for identifying a circuit comprising a filter in series with a spiking neuron model, based only on a relatively small number of input-output measurements, assuming that no input measurements of the neuron are available, and that the neuron parameters and the structure of the filter are unknown *a priori*. Therefore, the new approaches eliminate a number of assumptions of the previous methods.

Both approaches involve the estimation of the spiking neuron parameters first followed by the identification of the linear or nonlinear filter. A new technique is introduced that estimates the spiking neuron parameters using only the responses of the circuit to specific stimulus sequences. In both cases, the convergence to the true neuron parameters is guaranteed by proposed theoretical results, and practical algorithms are given to estimate the parameters in a realistic noisy environment.

The first approach addresses the problem of identifying a NF-IIF circuit. The parameters of the spiking neuron are estimated first, which allows reconstructing the nonlinear filter output (the IIF input) from the NF-IIF circuit output. Subsequently, the NARMAX methodology is applied to perform structure detection and parameter estimation of the nonlinear filter based on the input and the reconstructed filter output. The NARMAX methodology is arguably the most complete and advanced nonlinear system identification methodology, covering all aspects from stimulus design to model selection, parameter estimation and model validation (Billings & Chen, 1989; Billings et al.,

1988, 1989; Billings, 2013). This methodology has been successfully applied to characterize, directly or indirectly, neural processing circuits (Coca et al., 2000b; Friederich et al., 2016; Wei et al., 2009).

The second approach addresses the problem of identifying a LF-LIF circuit. A new algorithm is developed for estimating the LIF model parameters and, subsequently, the NARMAX methodology is used to infer the structure and estimate the parameters of the filter.

This paper is structured as follows. Section 2 introduces the proposed NF-IIF circuit model, and presents new theoretical results that enable redefining the identification problem for a circuit with fewer parameters, in two steps: the identification of the spiking neuron and the identification of the nonlinear filter. Section 3 introduces a new identification method for LF-LIF circuits. The conclusions are in Section 4.

2 A new method for identifying NF-IIF circuits from spike time sequences

The proposed circuit consists of a nonlinear filter connected in series with an IIF neuron, as depicted in Figure 1.

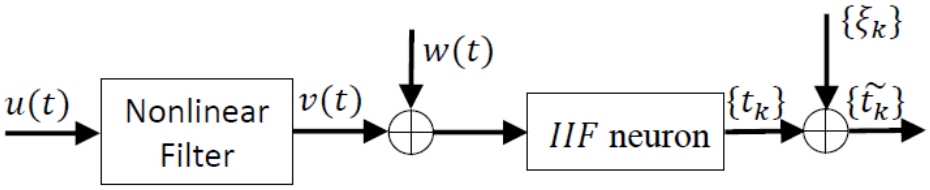

Figure 1: The structure of the circuit proposed for identification.

The nonlinear filter is described by the following equations

$$\begin{cases} \frac{dx}{dt}(t) = h_1\left(x(t), u(t)\right), \\ \\ v(t) = h_2\left(x(t), u(t)\right), \end{cases} \tag{1}$$

where $h_1 : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $h_2 : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ are nonlinear functions, $u(t)$ and $v(t)$ are the filter input and output, respectively, and $x : \mathbb{R} \to \mathbb{R}^n$ is the state variable vector. Let x_0 be the initial condition of system (1).

The system (1) is assumed to have an input-output representation

$$\frac{d^n v}{dt^n} = h\left(v, v', \ldots, v^{(n-1)}, u, u', \ldots, u^{(n_u-1)}\right),$$

where $1 \leq n_u \leq n$, i.e., the system is casual, and $h : \mathbb{R}^{n+n_u} \to \mathbb{R}$.

Let $v^0(t)$ be the response of the nonlinear filter to a step input $u^0(t) = A \cdot 1_{[0,\infty[}(t)$, $\forall t \in \mathbb{R}$, where $1_{[0,\infty[}(t)$ is the characteristic function of interval $[0, \infty[$. The filter is assumed to be bounded-input bounded-output (BIBO)-stable and that, $\forall A \in \mathbb{R}, v^0(t)$ converges to a steady state value v^0_∞, i.e., $\exists \lim_{t\to\infty} v^0(t) = v^0_\infty$. In other words, this assumes that the system is globally asymptotically stable, which is a reasonable assumption for the model of a sensory system (Smith, 2008). The filter output is assumed to be corrupted by Gaussian white noise $w(t)$ with zero mean and standard deviation σ_w.

The IIF neuron with capacitance C, threshold δ and bias b, denoted $IIF_{\{C,\delta,b\}}$, is described by the t-transform equation (Lazar & Pnevmatikakis, 2008)

$$\int_{t_k}^{t_{k+1}} v(\tau)d\tau = C\delta - b(t_{k+1} - t_k), \tag{2}$$

for $\forall k \in \mathbb{Z}$, where $v(t)$ is the neuron input.

The IIF input $v(t)$ can be perfectly reconstructed (Lazar & Pnevmatikakis, 2011) if $v(t) > -b, \forall t \in \mathbb{R}$, $v \in PW_\Omega$ and $\frac{b}{C\delta} > \frac{\Omega}{\pi}$, where PW_Ω is the Paley-Wiener space of bandwidth $\Omega > 0$

$$PW_\Omega = \left\{ v \in L^2(\mathbb{R}) : \text{supp}(\mathcal{F}v) \subseteq [-\Omega, \Omega] \right\},$$

where $\mathcal{F}v(j\omega)$ is the Fourier transform of $v(t)$ and supp$(\mathcal{F}v)$ denotes the support of $\mathcal{F}v(j\omega)$. For a function that is not bandlimited, or whose bandwidth is unknown, there are alternative reconstruction methods available (Lazar & Pnevmatikakis, 2010; Lazar et al., 2010).

In the following it is assumed that, for any $u(t)$, the output of the nonlinear filter (1) satisfies $v \in PW_\Omega$, such that $\Omega < \frac{\pi b}{C\delta}$.

The observed spike times sequence generated by the IIF neuron is assumed to be corrupted by uniform noise $\{\xi_k\}_{k\in\mathbb{Z}}$ with zero mean and amplitude A_ξ, which models the error associated with the measurement of the spike times $\{t_k\}_{k\in\mathbb{Z}}$.

2.1 An identification method based on an equivalent NF-IIF circuit

To simplify the identification problem, an equivalent model of the NF-IIF circuit, which involves a single tunable parameter, is derived first. This strategy was used in (Lazar & Slutskiy, 2010) for identifying the spiking neuron component of a LF-IIF circuit. Here, this approach is extended to NF-IIF circuits.

Two NF-IIF circuits are said to be input-output equivalent if, given input function $u(t)$, they generate the same output spike times $\{t_k\}_{k\in\mathbb{Z}}$. The equivalence relation is a consequence of the following lemma.

Lemma 1. *Let $\{t_k\}_{k\in\mathbb{Z}}$ be the sequence of spike times generated by neuron $IIF_{\{C,\delta,b\}}$*

given input $v(t)$*. Let* r *be an arbitrary number satisfying* $r > -b$*. Then the following holds true*

$$\int_{t_k}^{t_{k+1}} y(\tau)d\tau = \delta_b - (t_{k+1} - t_k), \forall k \in \mathbb{Z},$$

where $\delta_b = \frac{C\delta}{b+r}$ *and* $y(t) = \frac{v(t)-r}{b+r}, \forall t \in \mathbb{R}$.

Proof. The t-transform of $IIF_{\{C,\delta,b\}}$ satisfies (2)

$$\int_{t_k}^{t_{k+1}} v(\tau)d\tau = C\delta - b(t_{k+1} - t_k) = C\delta - (b+r)(t_{k+1} - t_k) + r(t_{k+1} - t_k)$$

$$\Leftrightarrow \int_{t_k}^{t_{k+1}} (v(\tau) - r)\, d\tau = C\delta - (b+r)(t_{k+1} - t_k). \tag{3}$$

The required result follows after dividing both sides of (3) by $(b + r)$. □

In essence, the previous result demonstrates that the neuron $IIF_{\{C,\delta,b\}}$ with input $v(t)$ generates the same spike times $\{t_k\}_{k\in\mathbb{Z}}$ as the neuron $IIF_{\{1,\delta_b,1\}}$ with input $y(t)$.

In practice, $r = r(u)$ is the steady state output of the nonlinear filter in response to a step input. As a consequence, it follows that the NF-IIF circuits depicted in Figure 1 and Figure 2 are input-output equivalent.

A method to identify the circuit in Figure 2, which involves first the identification of the spiking neuron followed by the identification of the nonlinear filter, is summarized below.

Step 1. Spiking neuron parameter estimation

For a given filter input $u^0(t) = A \cdot 1_{[0,\infty[}(t), \forall t \in \mathbb{R}$, it is assumed that the output of the NF-IIF circuit is $\{t_k^0\}_{k\in\mathbb{Z}}$, which corresponds to the nonlinear filter output $v^0(t)$.

It is assumed that the filter input amplitude A is selected such that $v_\infty^0 = \lim_{t\to\infty} v^0(t) > -b$. In essence, this means that the IIF neuron generates spikes in response to a step

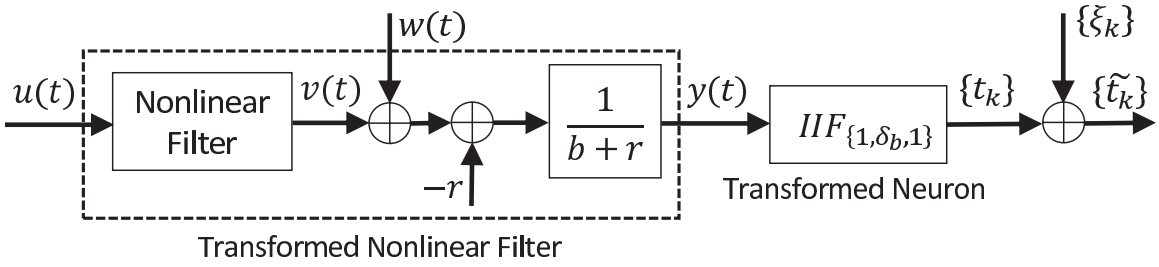

Figure 2: Input-output equivalent NF-IIF circuit.

input of amplitude v_∞^0. According to Lemma 1 for $v(t) = v^0(t)$ and $r = v_\infty^0$, in the absence of noise, it follows that $\lim_{k\to\infty} \int_{t_k}^{t_{k+1}} y(\tau)d\tau = 0$, and thus

$$\lim_{k\to\infty} \Delta t_k^0 = \delta_b,$$

where $\Delta t_k^0 = t_{k+1}^0 - t_k^0, \forall k \in \mathbb{Z}$.

In a more realistic scenario assuming the presence of noise and that only a finite number of noise corrupted spike times $\{\widetilde{t}_k^0\}_{k=1}^N$ are available, an estimate of the parameter δ_b is given by

$$\widehat{\delta}_b = \frac{\sum_{k=k_0}^{N-1} \widetilde{\Delta t_k^0}}{N - k_0}, \tag{4}$$

where k_0 satisfies $\left|\widetilde{\Delta t_k^0} - \frac{1}{N-k}\sum_{i=k}^{N-1} \widetilde{\Delta t_i^0}\right| < \Delta t_{err}^0, \forall k = k_0, \ldots, N$, and Δt_{err}^0 is a parameter selected by the user.

Step 2. Estimation and structure detection of the nonlinear filter

Let $\{\widetilde{t}_k\}_{k\in\mathbb{Z}}$ be the noisy output of the NF-IIF circuit given the input $u(t)$. The output $y(t)$ of the transformed nonlinear filter in Figure 2 is reconstructed from the spike times $\{\widetilde{t}_k\}_{k\in\mathbb{Z}}$, assuming that they are generated by the neuron $IIF_{\{1,\widehat{\delta}_b,1\}}$, where $\widehat{\delta}_b$ is the

estimated in the previous step. The reconstruction is performed with the algorithm introduced by Lazar & Pnevmatikakis (2010). This function reconstructed with this algorithm is consistent, i.e., it triggers the same spike times when encoded with the same IIF neuron and, additionally, it minimizes a smoothness criterion.

In practice $u(t)$ and $\hat{y}(t)$ are sampled with period ε_1, which is usually too small to enable the correct identification of the nonlinear filter. For this reason, the functions $u(t)$ and $\hat{y}(t)$ are then downsampled to period $\varepsilon_2 \geq \varepsilon_1$ before performing system identification. The value of ε_2 is selected using the procedure in (Billings & Aguirre, 1995), which is known to produce improved results for identification problems.

Let $u[k]$ and $\hat{y}[k]$ be the input and output sequences of the nonlinear filter, sampled with the period ε_2. Given the input/output data, the NARMAX system identification methodology is used to infer a NARMAX model (Leontaritis & Billings, 1981)

$$\hat{y}[k] = F(\hat{y}[k-1], \ldots, \hat{y}[k-n_y], u[k-1], \ldots, u[k-n_u], \\ e[k-1], \ldots, e[k-n_e]) + e[k],$$

where $e[k]$ represents the combined effects of measurement noise, modelling errors and unmeasured disturbances, n_u, n_y and n_e are constants denoting the maximum input, output and noise lags, respectively, and $F : \mathbb{R}^{n_y+n_u+n_e} \to \mathbb{R}$ is a multivariate polynomial of degree l. The structure and parameters are assumed to be unknown and are determined using the Orthogonal Forward Regression (OFR) algorithm (Chen et al., 1989). Specifically, given a set of candidate regressors consisting of all possible monomials $\{p_i\}_{i=1}^{M}$, $p_i : \mathbb{R}^{n_y+n_u+n_e} \to \mathbb{R}$, a greedy iterative selection algorithm is employed which, at each step, selects the regressor that contributes the most to the reduction of the error. The process terminates when the estimated model equation satisfies an infor-

mation theoretic criterion (Akaike, 1969). The resulting model is given by

$$\hat{y}[k] = \sum_{s=1}^{m} \theta_s p_s(X[k]) + e[k],$$

where

$$X[k] = [\hat{y}[k-1], \ldots, \hat{y}[k-n_y], u[k-1], \ldots, u[k-n_u],$$
$$e[k-1], \ldots, e[k-n_e]].$$

To validate the model we compute the model predictions for a stimulus function not used in identification and calculate the normalized mean squared error between the output reconstructed with the method in (Lazar & Pnevmatikakis, 2010) and the model predicted output (Billings, 2013)

$$NMSE = \frac{\|\hat{\hat{y}}[k] - \hat{y}[k]\|^2_{\ell^2}}{\|\bar{\hat{y}} - \hat{y}[k]\|^2_{\ell^2}},$$

where $\hat{\hat{y}}[k]$ is the model predicted output sequence, $\bar{\hat{y}}$ is the average of the sequence $\hat{y}[k]$, and $\|\cdot\|_{\ell^2}$ denotes the norm in space ℓ^2 .

To further evaluate the extend to which the identified nonlinear model captured the dynamic characteristics of the system, we compute and compare the Generalized Frequency Response Functions (GFRFs), of the original and identified model (Billings, 2013). The NARMAX model could also be mapped onto a continuous-time equivalent model, for example using an approach based on the GFRFs calculated for the NARMAX model (Swain et al., 1998), which would allow simulating the system at any desired sampling period.

2.2 Numerical study

The performance of the proposed identification method is demonstrated using four numerical examples.

In the first example, the nonlinear filter that satisfies the fading memory requirement, and thus can be represented as a Volterra series, is considered. The second example demonstrates the more general applicability of our approach, by considering a case where the spiking neuron does not satisfy the proposed assumptions. Specifically, the circuit consists of a nonlinear filter in series with a HH neuron (NF-HH), where the HH neuron is connected via multiplicative coupling (Lazar & Slutskiy, 2010). The third example considers a NF-IIF circuit where the dynamics of the nonlinear filter are chaotic and cannot be described by a Volterra series. cannot be described by a Volterra system. The fourth example tests the proposed methodology using input-output recordings from a spiking neuron located in the primary visual area of the mouse.

Example 1.

The nonlinear filter block of the NF-IIF circuit is described by the following equation

$$v''(t) + \alpha v'(t) + \beta v(t) + \gamma \left(v(t)\right)^2 = u(t), \forall t \in \mathbb{R}, \tag{5}$$

where $\alpha = 0.2, \beta = 1, \gamma = 0.1$. The output of the nonlinear filter is corrupted by additive Gaussian white noise $w(t)$ with zero mean and standard deviation $\sigma_w = 10^{-2}$. The nonlinear system is connected in series with an IIF neuron with parameters $b = 15, \delta = 3$ and $C = 1$. It is assumed that the output spike times sequence is noise free, i.e., $A_\xi = 0$.

Step 1. Spiking neuron parameter estimation

The NF-IIF was simulated numerically using a step input $u^0(t) = 1_{[0,\infty[}(t)$ with duration $T = 180\ s$, sampled with period $\varepsilon_1 = 10^{-2}\ s$. The selected value of T is longer than the transient regime of the nonlinear system response. The differential equation (5) was solved numerically to compute the nonlinear system output $v^0(t)$ using the $ode15s$ routine in Matlab with fixed time step ε_1.

The output spike train of the IIF neuron $\{t_k^0\}_{k=1}^N$, where $N = 995$, is computed as

$$t_k^0 = (l_k + 1)\varepsilon_1 - \varepsilon_1 \cdot \frac{U((l_k + 1)\varepsilon_1) - kC\delta}{U((l_k + 1)\varepsilon_1) - U(l_k\varepsilon_1)}, k = 1, \ldots, N, \tag{6}$$

where $U(l_k\varepsilon_1) = \int_0^{l_k\varepsilon_1}(u^0(\tau) + b)d\tau$ is computed using the trapezoid rule, ε_1 is the sampling time, and l_k is the unique solution of

$$U(l_k\varepsilon_1) \leq kC\delta < U((l_k + 1)\varepsilon_1).$$

The parameter δ_b was estimated for $k_0 = 125$ satisfying $\left|\Delta t_k^0 - \frac{1}{N-k}\sum_{i=k}^{N-1}\Delta t_i^0\right| < 10^{-3}, \forall k = k_0, \ldots, N$.

The constant v_∞^0 is estimated as $v_\infty^0 = v^0(180) = 0.916$. Given δ, b, C, and v_∞^0, δ_b was calculated as $\delta_b = \frac{C\delta}{b+v_\infty^0} = 0.1885$. In this particular case, the estimation error of δ_b was $e_{\delta_b} = \delta_b - \hat{\delta}_b = 5.92 \cdot 10^{-7}$.

Step 2. Estimation and structure detection of the nonlinear filter

The data used to identify the nonlinear filter was generated by simulating the NF-IIF circuit using an input function $u_{tr}(t)$. The sampling period was $\varepsilon_1 = 10^{-2}\ s$ and the duration $T = 180\ s$. The samples were drawn from $N(0, 1)$. The input is subsequently low-pass filtered to $\Omega_0 = 4\ rad/s$ using a Butterworth filter with bandpass corner

frequency $2\ rad/s$, stopband corner frequency $4\ rad/s$, maximum attentuation in the passband of $10\ dB$, and minimum attenuation in the stopband of $40\ dB$. The input was subsequently normalized such that $|u_{tr}(t)| \leq 1$.

The output of the circuit consisted of a spike time sequence $\{t_k^{tr}\}_{k=1}^{898}$. To validate the model, a separate circuit input $u_{val}(t)$ and output sequence $\{t_k^{val}\}_{k=1}^{897}$ were generated using the above procedure.

The output signal used to identify the filter was reconstructed first, based on the spike time sequence $\{t_k^{tr}\}_{k=1}^{898}$ and the spiking neuron model identified in step 1, and the sampling period is $\varepsilon_1 = 10^{-2}\ s$.

Functions $u_{tr}(t)$ and $\hat{y}_{tr}(t)$ were preprocessed to remove the mean. To ensure that the distortions of the reconstructed filter output due to boundary effects are not affecting the identification procedure, the first and last 1800 samples were discarded. The resulting functions are depicted in Figure 3.

The input/output data used to identify the transformed nonlinear filter was obtained by downsampling the original data sampled at ε_1. The sampling period used in identification $\varepsilon_2 = 0.15\ s$ was determined using the approach proposed by Billings & Aguirre (1995).

The input and output data used in identification $u_{tr}[k], \hat{y}_{tr}[k]$ was subsequently obtained by downsampling the original data.

The degree of nonlinearity and the maximum number of input and output lags to initialize the regression for the NARMAX model were determined iteratively starting from small values. The best results in terms of prediction performance and model size were $l = 2$, $n_u = 10, n_y = 10$, and $n_e = 0$. The OFR algorithm selected, in a stepwise

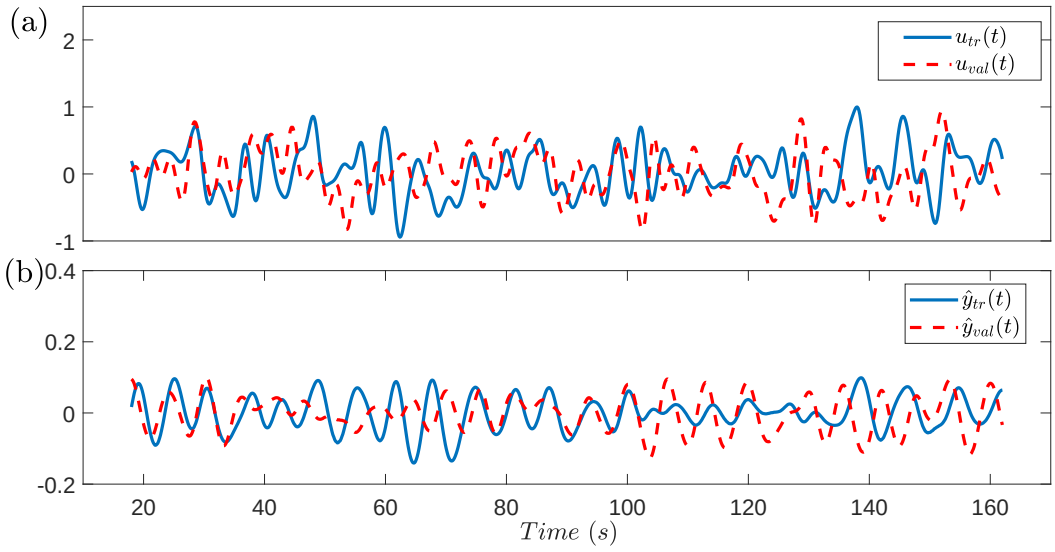

Figure 3: Filter identification and validation inputs and outputs: a) filter inputs and b) reconstructed filter outputs.

manner, an increasing number of regressors until the stop criterion $NMSE < 7 \cdot 10^{-4}$, was met for $m = 10$. The final set of regressors $\{p_s(X[k])\}_{s=1}^{m}$ and the corresponding estimated parameters $\{\theta_s\}_{s=1}^{m}$ are presented in the Appendix A.

The model predicted output $\hat{\hat{y}}_{val}[k]$, computed using the validation input $u_{val}[k]$, is shown in Figure 4a. The corresponding model prediction error $e_{val}[k]$ is shown in Figure 4b. The NMSE for estimation and validation are $2.52 \cdot 10^{-4}$ and $2 \cdot 10^{-4}$, respectively.

The magnitude functions for the first and second order GFRFs for the original system (5), derived in (Li & Billings, 2011), are given by

$$H_1(j\omega) = \frac{1}{-\omega^2 + \alpha j\omega + \beta},$$

$$H_2(j\omega_1, j\omega_2) = -\gamma H_1(j\omega_1) H_1(j\omega_2) H_1(j\omega_1 + j\omega_2).$$

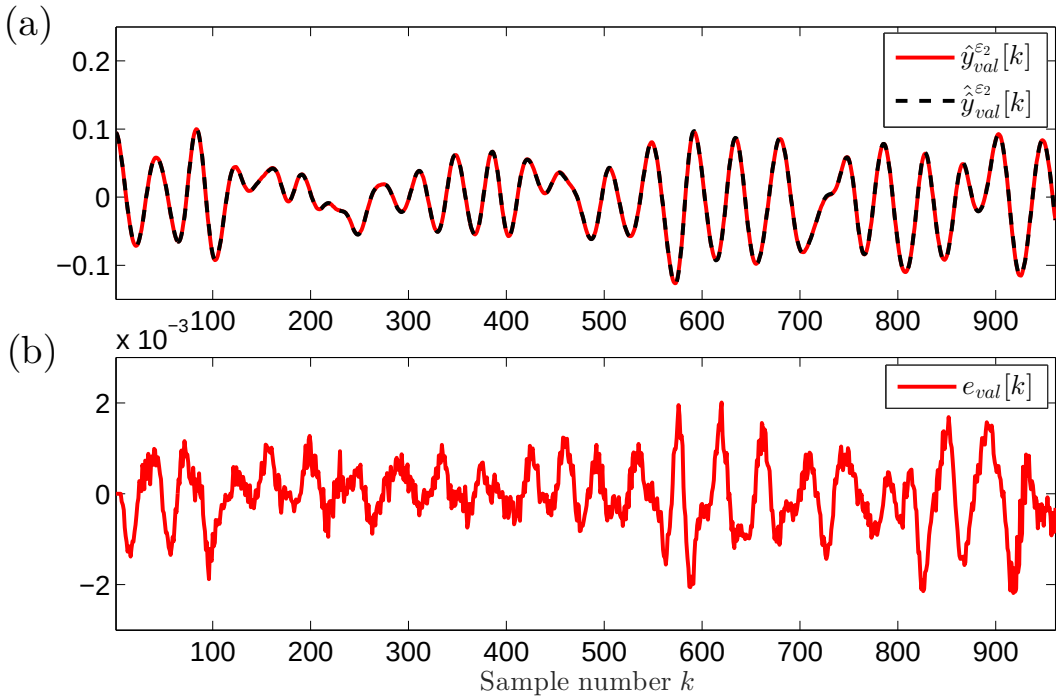

Figure 4: (a) Validation $\hat{y}_{val}[k]$ and the model predicted output $\hat{\hat{y}}_{val}[k]$. (b) The model predicted error $e_{val}[k]$.

The identified NARMAX model is used to derive analytically the first and second order generalized frequency response functions $\hat{H}_1(j\omega)$ and $\hat{H}_2(j\omega_1, j\omega_2)$ (Billings, 2013). The following errors are defined for quantifying the error between the GFRFs of the original and identified transformed filter

$$E_1(j\omega) = 100 \cdot \frac{|H_1(j\omega)| - \left|\hat{H}_1(j\omega)(b + v_\infty^0)\right|}{\|H_1\|_\infty}(\%), \tag{7}$$

$$E_2(j\omega_1, j\omega_2) = 100 \cdot \frac{|H_2(j\omega_1, j\omega_2)| - \left|\hat{H}_2(j\omega_1, j\omega_2)(b + v_\infty^0)\right|}{\|H_2\|_\infty}(\%), \tag{8}$$

where $\|H_1\|_\infty = \max_{\omega\in\mathbb{R}} |H_1(j\omega)|$ and $\|H_2\|_\infty = \max_{\omega_1,\omega_2\in\mathbb{R}} |H_2(j\omega_1, j\omega_2)|$.

The functions $H_1(j\omega)$ and $H_2(j\omega_2, j\omega_2)$ are shown in Figure 5, and the error functions $E_1(j\omega)$ and $E_2(j\omega_1, j\omega_2)$ are illustrated in Figure 6.

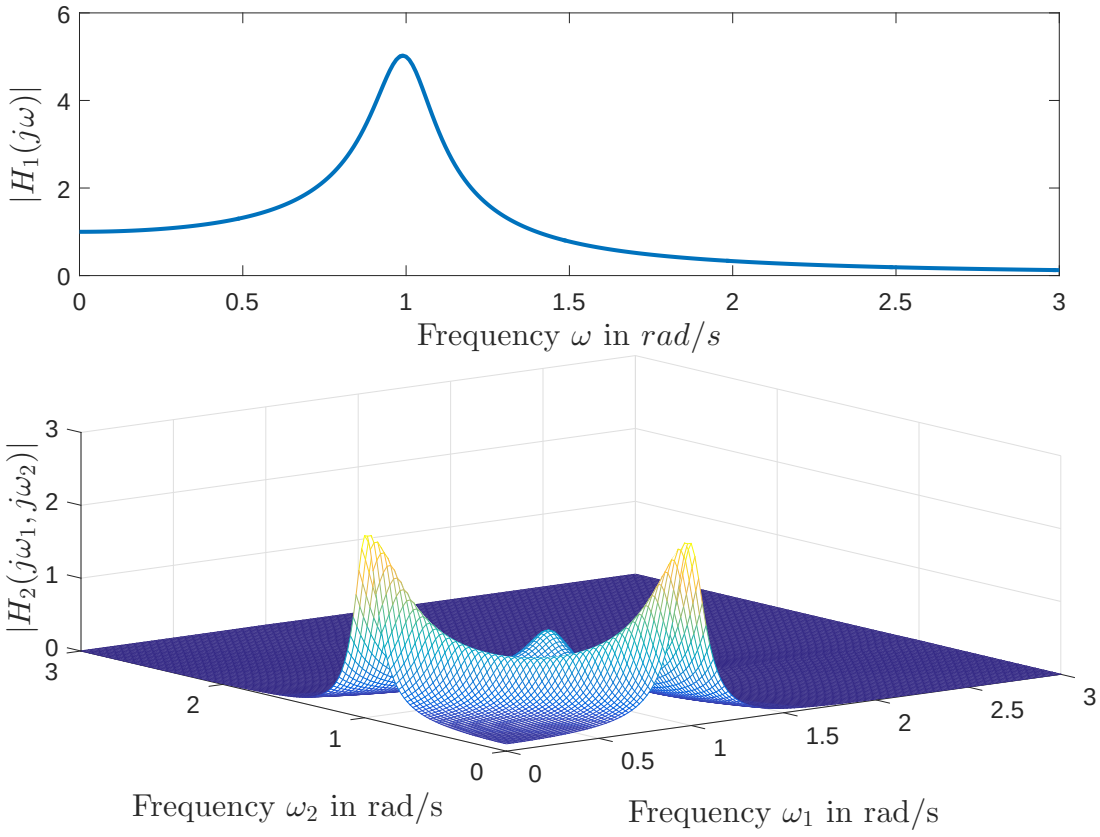

Figure 5: The absolute values of the GFRF functions $H_1(j\omega)$ and $H_2(j\omega_1, j\omega_2)$, associated with system (5).

The NARX model was inferred from input/output measurements sampled with period $\varepsilon_2 = 15 \cdot \varepsilon_1$, which is often too large for computing accurately the output spike times of the NF-IIF circuit. In order to simulate the circuit with inputs $u_{tr}(t), u_{val}(t)$, sampled with period ε_1, a new set of inputs $u^i_{tr}[k], u^i_{val}[k], i = 1, \ldots, 15$, were gener-

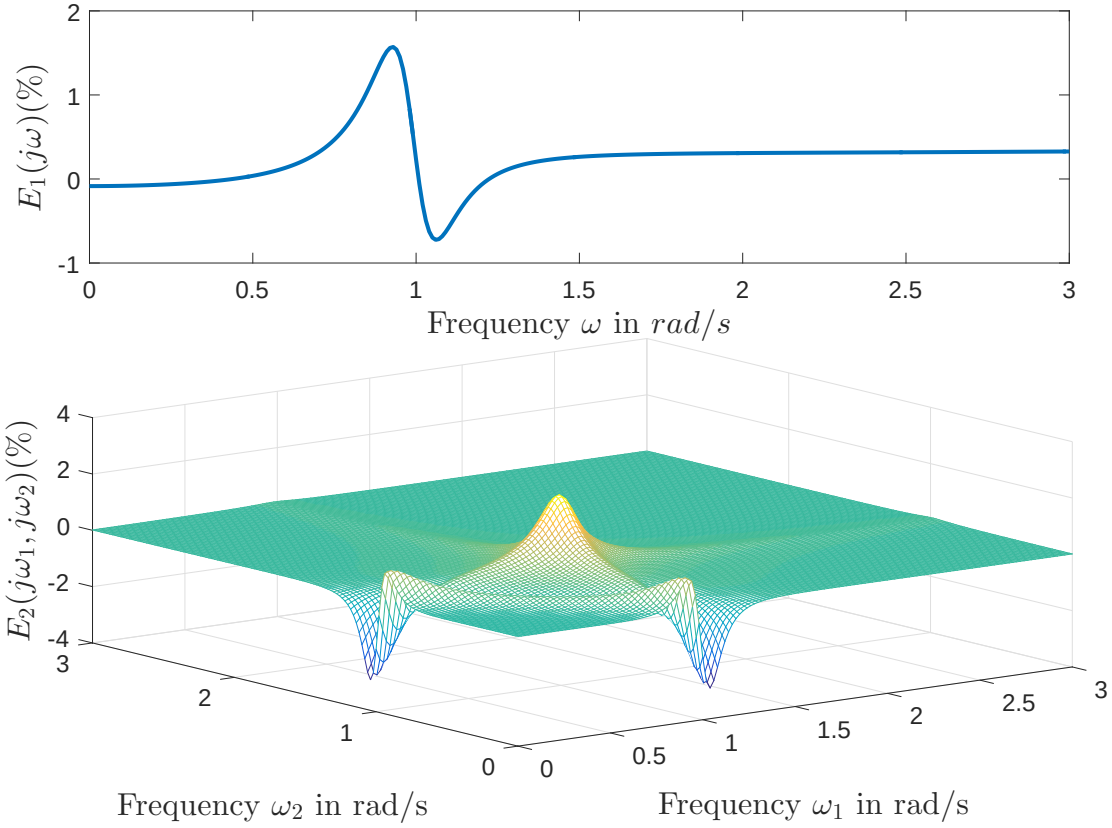

Figure 6: The error functions $E_1(j\omega)$ and $E_2(j\omega_2, j\omega_2)$.

ated, satisfying

$$\begin{aligned} u_{tr}^{i}[k] &= u_{tr}\left((i+15k)\varepsilon_1\right), \\ u_{val}^{i}[k] &= u_{val}\left((i+15k)\varepsilon_1\right), i = 1, \ldots, 14. \end{aligned} \tag{9}$$

Essentially, the filter inputs $u_{tr}^{i}[k], u_{val}^{i}[k]$ represent the samples of $u(t)$ measured with period ε_2, where the first sampling time is $i\varepsilon_1$, respectively. The output of the filter for the required sampling time can then be computed by simulating the NARX model with these inputs, for every i, as follows.

The corresponding outputs of the NARX system, given the inputs above, consisted

of $\hat{\hat{y}}^i_{tr}[k], \hat{\hat{y}}^i_{val}[k], i = 1, \ldots, 15$. The functions $\hat{\hat{y}}_{tr}(t)$ and $\hat{\hat{y}}_{val}(t)$, sampled with ε_1, were computed as

$$\begin{aligned} &\hat{\hat{y}}_{tr}\left((15k+i)\varepsilon_1\right) = \hat{\hat{y}}^i_{tr}[k], \\ &\hat{\hat{y}}_{val}\left((15k+i)\varepsilon_1\right) = \hat{\hat{y}}^i_{val}[k], i = 1, \ldots, 14. \end{aligned} \tag{10}$$

Finally, the neuron $IIF_{\{1,\hat{\delta}_b,1\}}$ generated spike time sequences $\{\hat{t}^{tr}_k\}_{k=1}^{719}$ and $\{\hat{t}^{val}_k\}_{k=1}^{718}$ in response to inputs $\hat{\hat{y}}_{tr}(t), \hat{\hat{y}}_{val}(t)$, respectively.

The rate of coincidence between two sequences of spike times was evaluated by computing the coincidence factor Γ, introduced by (Jolivet et al., 2006) where

$$\Gamma = \frac{N_{coinc} - \langle N_{coinc} \rangle}{0.5(N_{data} + N_{model})} \frac{1}{\mathcal{N}},$$

where N_{data} is the number of spikes in the reference spike train, N_{model} is the number of spikes predicted by the NF-IIF model, N_{coinc} is the number of coincidences with precision Δ between the two spike trains, $\langle N_{coinc} \rangle = 2N_{model}\Delta N_{data}\frac{1}{T}$ is the expected number of coincidences by chance, and $\mathcal{N} = 1 - 2N_{model}\Delta\frac{1}{T}$, where T denotes the time duration of the simulation. The coincidence factor satisfies $\Gamma = 1$ only when there is complete coincidence with precision Δ between the predicted and the reference spike train, respectively. Moreover, a homogeneous Poisson process with a rate equal to $N_{model}\frac{1}{T}$ has a coincidence factor $\Gamma = 0$. The exact value for Δ is not critical and, for experimental data, Jolivet et al. (2006) introduce the constraint $\Delta \in [1\ ms, 4\ ms]$. For the synthetic data used in this example, we selected $\Delta = 0.025\ s$, which satisfies $\Delta << 0.5 \cdot \min_k(t^{tr}_{k+1} - t^{tr}_k) = 0.09\ s$.

In this example, the coincidence factor was $\Gamma_{tr} = 1$ for the training data and $\Gamma_{val} = 1$ for the validation data. The values correspond to a percentage of correctly predicted spike times of 100%.

To evaluate the effect of noise on the identification method, the procedure was carried out for different levels of noise applied to the filter output and the measurement of the spike times. The results in Table 1 show that e_{δ_b} is not changing significantly for different noise levels. Because the NMSE errors are higher when the spike times are corrupted by noise, i.e. $A_\xi \neq 0$, the number of regressors was increased in this case to $m = 7$.

Table 1: The identification results in Example 1 for different values of σ_w and A_ξ.

σ_w	A_ξ	e_{δ_b}	m	$NMSE_{tr}$	$NMSE_{val}$	Γ_{tr}	Γ_{val}
0.01	0	$6.92 \cdot 10^{-7}$	6	$8.45 \cdot 10^{-4}$	$5.3 \cdot 10^{-4}$	1	1
0.03	0	$9.51 \cdot 10^{-7}$	6	0.005	0.006	0.61	0.95
0.05	0	$1.21 \cdot 10^{-6}$	6	0.005	0.006	0.59	0.92
0.01	$3 \cdot 10^{-4}$	$1 \cdot 10^{-6}$	7	0.01	0.011	0.4	0.42
0.03	$3 \cdot 10^{-4}$	$1.26 \cdot 10^{-6}$	7	0.008	0.009	0.32	0.39
0.05	$3 \cdot 10^{-4}$	$1.52 \cdot 10^{-6}$	7	0.008	0.009	0.33	0.38

Example 2.

This example demonstrates that the proposed approach can be applied to identify a more biophysically realistic neural circuit, that does not satisfy the proposed assumptions. Specifically, the spiking neuron is represented as a Hodgkin-Huxley (HH) model, given by

$$\begin{aligned}
C\frac{dV}{dt} =& - g_{Na}m^3h(V - E_{Na}) - g_K n^4(V - E_K) - g_L(V - E_L) + I_b \\
\frac{dm}{dt} =& \alpha_m(V)(1 - m) - \beta_m(V)m \\
\frac{dh}{dt} =& \alpha_H(V)(1 - h) - \beta_h(V)h \\
\frac{dn}{dt} =& \alpha_n(V)(1 - n) - \beta_n(V)n,
\end{aligned}$$

where V is the membrane voltage of the neuron, m, h, n are the gating variables, and I_b is the injected current. The explicit values for each parameter can be found in (Izhikevich, 2007). Here, the value for the injected current was chosen $I_b = 120\ \mu A/cm^2$. The HH equations above can be rewritten as $\frac{dz}{dt} = f(z)$, where $z = [V, m, h, n]$ and $f : \mathbb{R}^4 \to \mathbb{R}^4$.

The proposed circuit consists of a nonlinear filter, described by system (5), connected via multiplicative coupling to a HH model, such that (Lazar & Slutskiy, 2010)

$$\frac{dz}{dt} = (b + v(t))f(z), \tag{11}$$

where b is a bias parameter. The output spike times $\{t_k\}_{k\in\mathbb{Z}}$ are defined as the local maxima of the voltage trace $z_1(t) = V(t)$, such that

$$\frac{dz_1}{dt}(t_k) = 0, \frac{d^2z_1}{dt^2}(t_k) < 0, \forall k \in \mathbb{Z}.$$

Lazar & Slutskiy (2010) have proven that the spiking neuron defined above is input-output equivalent to the neuron model $IIF_{\{1,\delta,b\}}$, where the δ depends on the HH parameters. The new proposed methodology is used in the following for identifying an input-output equivalent NF-IIF model for the proposed NF-HH circuit.

Step 1. Spiking neuron parameter estimation

The NF-HH circuit was excited with the same step input $u^0(t)$ as in Example 1. The output of the filter was $v^0(t)$ and the solution $z(t)$ of system (11) was computed using the $ode15s$ routine in Matlab with fixed step ε_1. The sequence of spike times $\{t_k^0\}_{k=1}^{450}$ was computed as the local maxima of $z_1(t)$.

The IF parameter δ_b was estimated for $k_0 = 150$ as $\hat{\delta_b} = 0.399245$.

Step 2. Estimation and structure detection of the nonlinear filter

The identification of the nonlinear filter is carried out as in Example 1. The final set of model terms selected for a stopping criterion $NMSE < 7 \cdot 10^{-4}$ is summarized in Table 5.

The NMSEs calculated for the training and validation data sets are $5.82 \cdot 10^{-5}$ and $1.35 \cdot 10^{-4}$, respectively. The error functions $E_1(j\omega)$ and $E_2(j\omega)$, computed between the magnitudes of the GFRFs, are shown in Figure 7.

The predicted spike time sequences generated by the identified NF-IIF circuit in response to the inputs $u_{tr}(t)$ and $u_{val}(t)$ were also compared with the original ones generated by the original NF-HH model. The coincidence factors for training and validation were $\Gamma_{tr} = 1$ and $\Gamma_{val} = 1$ respectively, corresponding to 100% correctly predicted spike times.

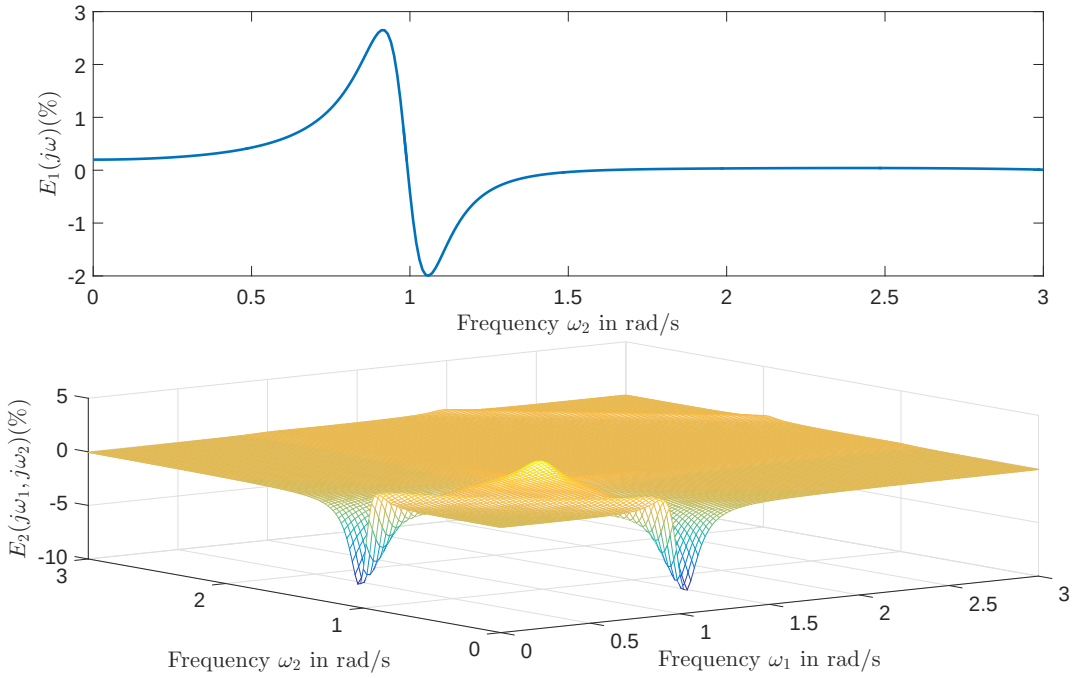

Figure 7: The error functions $E_1(j\omega)$ and $E_2(j\omega_2, j\omega_2)$.

Example 3.

In this example, the nonlinear filter block in Figure 1 is the well known Duffing-Ueda chaotic nonlinear dynamical system (Ueda, 1985)

$$v''(t) + kv'(t) + (v(t))^3 = u(t), \forall t \in \mathbb{R}, \tag{12}$$

where $k = 0.1$. The nonlinear system is connected in series with an IIF neuron with parameters $b = 15, \delta = 1.5$ and $C = 1$. In this example it is assumed that $\sigma_w = A_\xi = 0$.

The system (12) is solved using the $ode45$ Matlab routine, with initial conditions $v(0) = v'(0) = 0$. The output of the IIF neuron $\{t_k\}_{k=1}^N$ is computed with (6).

Step 1. Spiking neuron parameter estimation

To estimate the spiking neuron parameter, the response of the circuit to a step input

$u^0(t) = 1_{[0,\infty[}(t), \forall t \in [0,T], T = 720\ s$, was computed using a sampling period $\varepsilon_1 = \frac{\pi}{300}$. The output of the circuit consisted of a spike time sequence $\{t_k^0\}_{k=1}^N$, where $N = 7678$. The noise-free data was used to estimate $\hat{\delta}_b = \delta_b = \Delta t_{N-1}^0 = 9.375 \cdot 10^{-2}$.

Step 2. Estimation and structure detection of the nonlinear filter

An input function $u_{tr}(t) = 11 \cdot \cos(t)$ was generated, with sampling time ε_1 and duration $360\ s$. The spike time sequence generated by the NF-IIF circuit in response to input $u_{tr}(t)$ was $\{t_k^{tr}\}_{k=1}^{3582}$.

Function $\hat{y}_{tr}(t)$ was reconstructed from $\{t_k^{tr}\}_{k=1}^{3582}$. The functions $u_{tr}(t)$ and $\hat{y}_{tr}(t)$ are depicted in Figure 8.

The sampling period for identification was $\varepsilon_2 = \frac{\pi}{60}\ s$ and the original input and output data were downsampled appropriately to generate the data set used for identification.

The model was estimated from a set of 1771 candidate regressors corresponding to $l = 3$, $n_u = 10, n_y = 10$, and $n_e = 0$. The model terms selection and parameter estimation was performed using a final set of $m = 23$ regressors, input $u_{tr}[k]$, and output $\hat{y}_{tr}[k]$.

The selected model terms and parameters estimates corresponding to the identified NARMAX model are presented in Appendix A, Table 4.

It is well known that chaotic systems exhibit sensitivity to the initial conditions and thus validating them using the NMSE lacks consistency (Billings & Aguirre, 1993). Moreover, the chaotic response doesn't admit a Volterra series expansion, and thus cannot be validated by computing error functions (7), (8). The bifurcation diagram was proven to be a useful tool for assessing the characteristics of a system by revealing

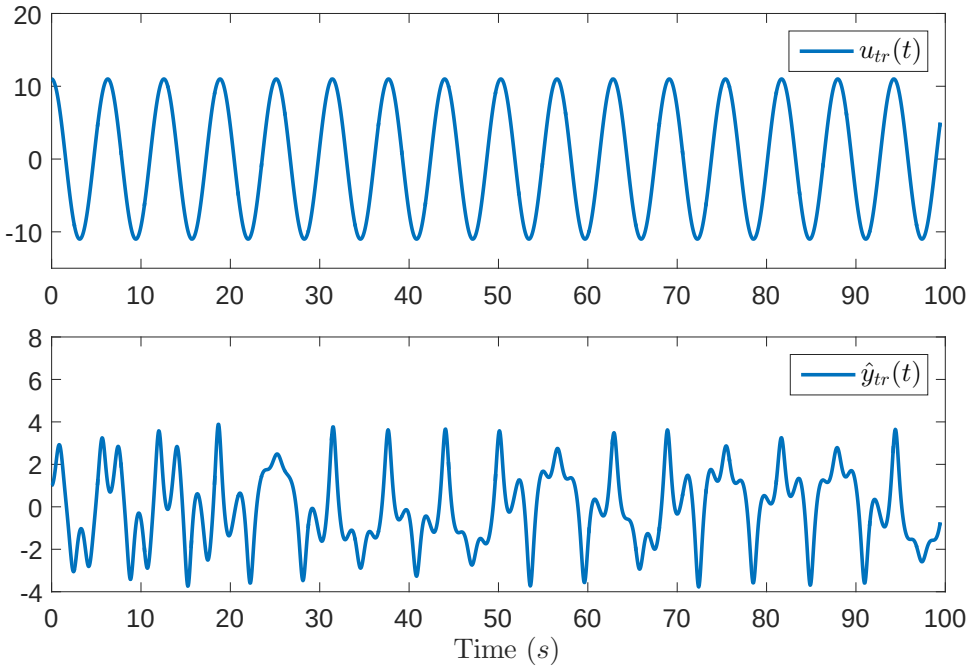

Figure 8: Filter input $u_{tr}(t)$ and the corresponding reconstructed nonlinear filter output $\hat{y}_{tr}(t)$.

at which values A it bifurcates, and also by detecting the parameter ranges for which the system shows chaotic behaviour (Billings & Aguirre, 1993).

The bifurcation diagrams of the true and identified nonlinear model, computed as in (Billings & Aguirre, 1993), are depicted in Figure 9.

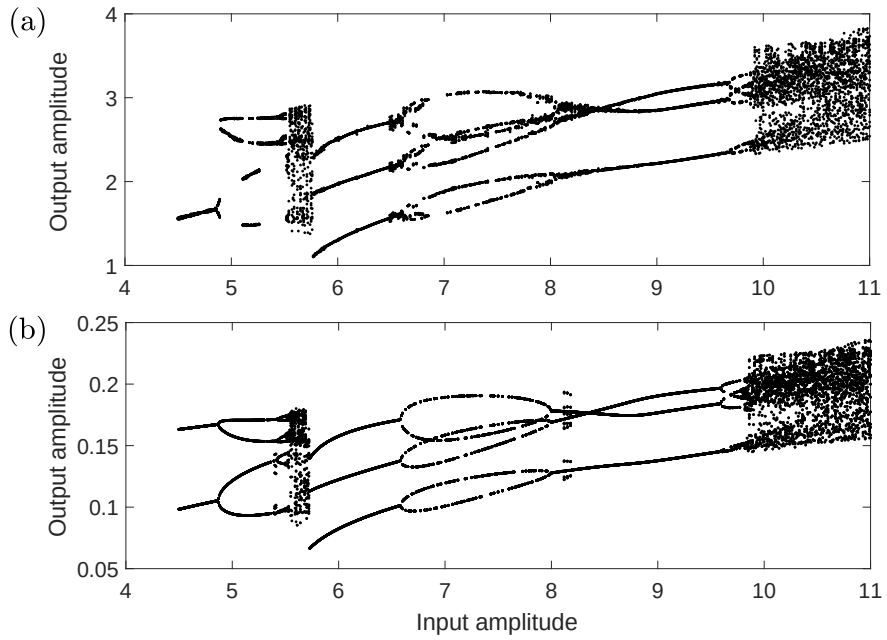

Figure 9: Bifurcation diagrams computed for the (a) original and (b) identified nonlinear filter with respect to the input amplitude A, where $u(t) = A \cdot \cos(t)$.

Example 4.

The proposed methodology is tested here using input-output recordings from a neuron located in the primary visual area of the mouse, layer 5. The data, recorded using brain slice electrophysiology, was downloaded from the Allen Cell Type Database (Allen Institute for Brain Science, 2015). The neuron selected has adaptation index 0.002, rheobase 390 pA, membrane time constant 7.7 ms, and firing rate 179.3 spikes/s (Allen Institute for Brain Science, 2016). Although the database provides recordings of the full voltage trace in response to stimuli, here only the spike times, computed as the peak values of the voltage trace, were used in the identification procedure.

Step 1. Spiking neuron parameter estimation

To estimate the IIF parameter, we used the response of the neuron to a long square stimulus with amplitude of 470 pA. The output spike times computed from the voltage trace are $\{t_k^0\}_{k=1}^N, N = 238$. The neuron parameter was estimated as $\hat{\delta}_b = 0.0044$ for $k_0 = 72$, which satisfies $|\Delta t_k^0 - \frac{1}{N-k}\sum_{i=k}^{N-1}\Delta t_i^0| < 2.2 \cdot 10^{-3}, \forall k > k_0$.

Step 2. Estimation and structure detection of the nonlinear filter

Two different periodic stimuli of duration $1\ s$ were used for the training and validation of the model, denoted $u_{tr}(t)$ and $u_{val}(t)$, consisting of pink noise with sampling rate $200\ kHz$, coefficient of variation of 0.2, amplitude of $555\ pA$, and period $1\ s$. The stimuli $u_{tr}(t), u_{val}(t)$ and the corresponding voltage traces $V_{tr}(t), V_{val}(t)$, recorded from the neuron, are depicted in Figure 10. The spike times used for the training and validation of the filter, computed from the voltage traces, are denoted $\{t_k^{tr}\}_{k=1}^{N_{tr}}\}$ and $\{t_k^{val,1}\}_{k=1}^{N_{val}}\}$, respectively, where $N_{tr} = 275, N_{val} = 277$.

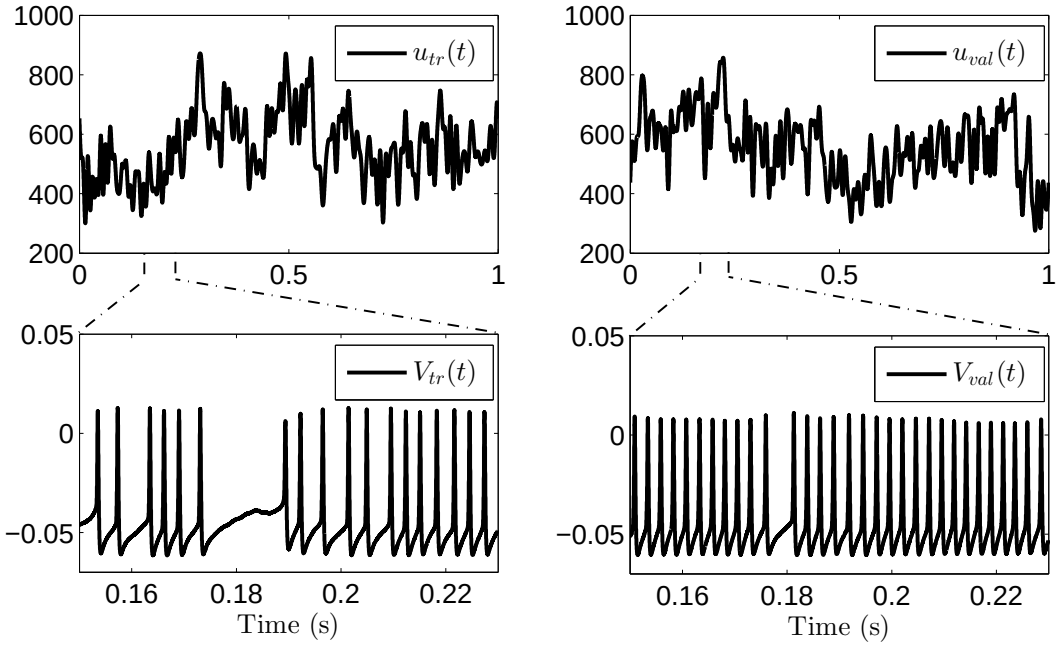

Figure 10: Stimuli $u_{tr}(t), u_{val}(t)$ and recorded neuron voltage traces $V_{tr}(t), V_{val}(t)$.

Given that the stimulus is periodic, in this example, the output of the nonlinear filter (input to the IIF neuron) was reconstructed using the algorithm proposed by Lazar et al. (2010), which uses a regularization parameter λ to trade off the consistency of the reconstruction, i.e., its ability to match the original spike times when encoded with the same neuron, for increased smoothness. Given that the output of a biological neuron is known to be highly corrupted by noise, this algorithm was found to give good results for reconstructing the nonlinear filter output. After a line search algorithm, the value $\lambda = 10^{-7}$ was found to lead to the smallest model predicted NMSE.

The filter output signals used for training and validation were reconstructed based on spike trains $\{t_k^{tr}\}_{k=1}^{N_{tr}}$ and $\{t_k^{val}\}_{k=1}^{N_{val,1}}$, respectively. The inputs and reconstructed outputs of the nonlinear filter $u_{tr}(t)$, $u_{val,1}(t)$, $\hat{y}_{tr}(t)$, $\hat{y}_{val,1}(t)$ are depicted in Figure 11.

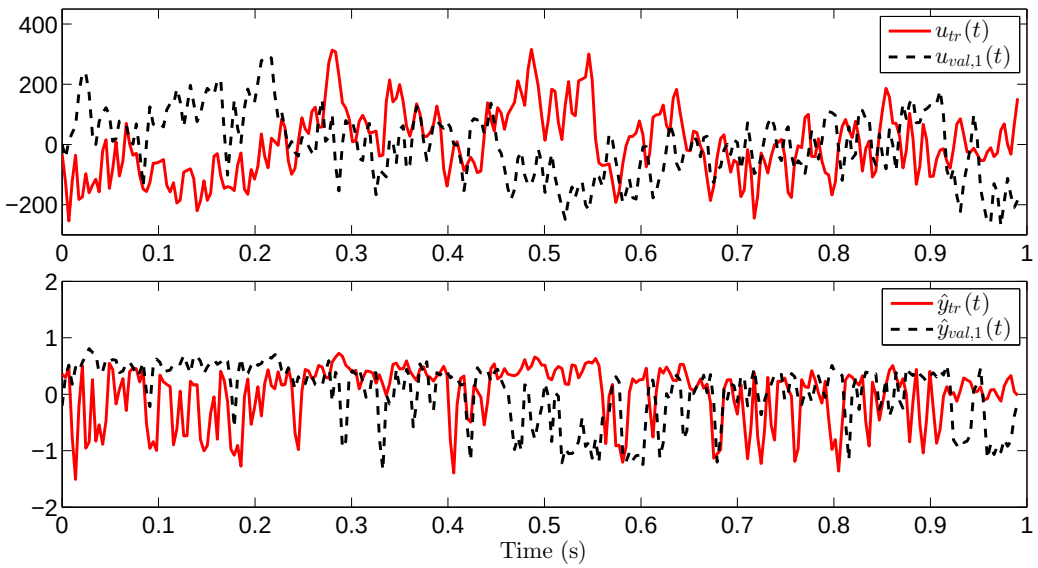

Figure 11: Filter input functions $u_{tr}(t), u_{val,1}(t)$ and the corresponding filter responses $\hat{y}_{tr}(t), \hat{y}_{val}(t)$, reconstructed using the method in (Lazar et al., 2010) with $\lambda = 10^{-7}$.

The data was subsequently downsampled and processed to remove the mean. The new sampling period is $\varepsilon_2 = 3.5 \cdot 10^{-3}\ s = 700 \cdot \varepsilon_1$.

The model was estimated from a set of 231 candidate regressor terms corresponding to $l = 2, n_u = 10, n_y = 10$, and $n_e = 0$. The OFR algorithm met the stop criterion $NMSE < 0.23$ for $m = 4$. The NARMAX model identified form the training data set is summarized in Appendix A.

The model predicted output $\hat{\hat{y}}_{val}[k]$, computed using the validation input $u_{val}[k]$, is shown in Figure 12a, superimposed over the output of the filter reconstructed using the original spike-time sequence. The corresponding model prediction error $e_{val}[k]$ is shown in Figure 12b. The NMSE for estimation and validation are 0.21 and 0.18, respectively.

The identified NF-IIF circuit was validated as before in terms of the output spike

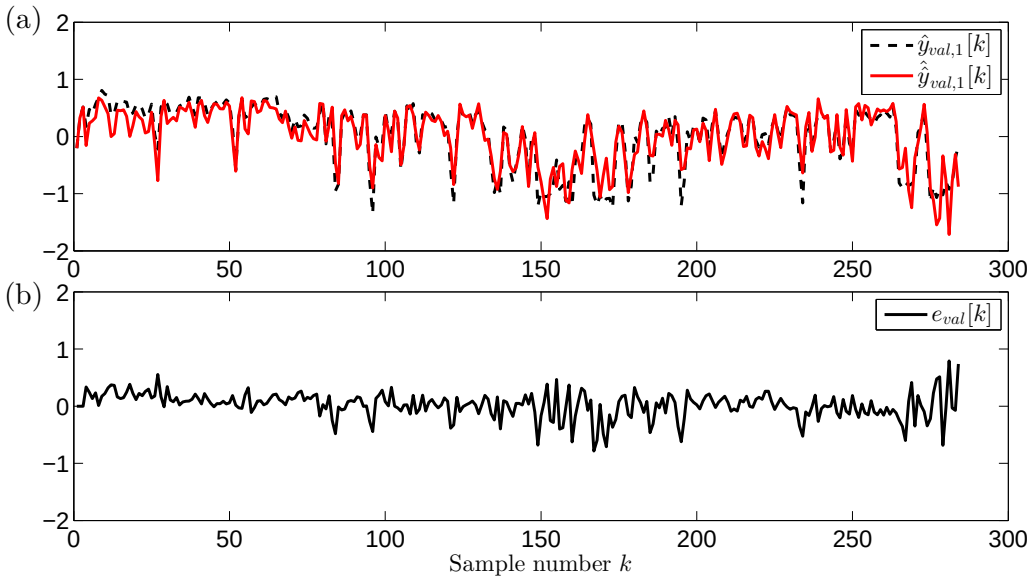

Figure 12: (a) The model predicted output $\hat{y}_{val}[k]$ superimposed on the reconstructed filter output reserved for model validation $\hat{\hat{y}}_{val,1}[k]$. (b) The model predicted error $e_{val}[k]$, computed over the validation dataset .

times prediction. For a precision of $\Delta = 1.5\ ms$, the coincidence factors for t_k^{tr} and t_k^{val} are $\Gamma_{tr} = 0.55$ and $\Gamma_{val} = 0.48$, respectively. The corresponding percentage of correctly predicted spike times is 91.9% and 92%, respectively. These performance indicators are in line with similar identification results for real data using simple threshold models (Jolivet et al., 2006). Although previous work has motivated the approximation of the subthreshold dynamics of the neuron under random current injection by a linear filter (Jolivet et al., 2006), this example gives more insight into these dynamics, by showing they have a significant nonlinear behaviour. This can be quantified in the proposed model by the ERR value of the nonlinear regressor, i.e., its percentage contribution to the model output, which amounts to 5.43% (see Appendix A). The proposed work also

has the advantage that it requires only extracellular recordings of the neuron (the spike times) unlike the method by Jolivet et al. (2006), that uses intracellular recordings for the fitting procedure (the whole voltage trace) .

3 A new method for identifying LF-LIF circuits from spike time sequences

The LF-LIF circuit (Figure 13) consists of a linear filter in series with a LIF neuron.

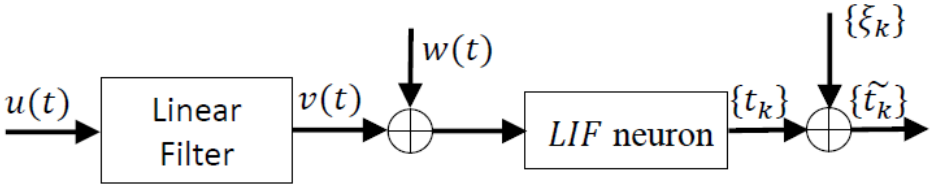

Figure 13: The structure of the circuit proposed for identification.

The linear filter has an impulse response function $g(t)$ satisfying $\int_{\mathbb{R}} |g(\tau)|d\tau < \infty$, that is, the filter is BIBO-stable. The filter gain K satisfies $K = \lim_{s\to\infty} G(s)$, where $G(s)$ denotes the Laplace transform of $g(t)$. It is assumed that the filter output is corrupted by Gaussian white noise $w(t)$ with zero mean and standard deviation σ_w.

The $LIF_{\{R,C,\delta,b\}}$ neuron is described by the t - transform equation (Lazar, 2005)

$$\int_{t_k}^{t_{k+1}} v(\tau)e^{-\frac{t_{k+1}-\tau}{RC}}d\tau = C(\delta - bR) + bRC \cdot e^{-\frac{t_{k+1}-t_k}{RC}}, \forall k \in \mathbb{Z}, \qquad (13)$$

where $v(t)$ is the neuron input, $\{t_k\}_{k\in\mathbb{Z}}$ denotes the spike time sequence generated by the LIF neuron, b is the bias, δ is the threshold, and R and C are the neuron resistance and capacitance, respectively.

Lazar (2005) has proven that the neuron input $v(t)$ can be reconstructed from the corresponding output spike time sequence $\{t_k\}_{k\in\mathbb{Z}}$ if $v \in PW_\Omega$ and

$$RC \cdot \ln\left(1 - \frac{\delta}{\delta - (b-c)R}\right)\frac{\Omega}{\pi} < \frac{1-\epsilon}{1+\epsilon},$$

$$v(t) > \frac{\delta}{R} - b, \forall t \in \mathbb{R}, \tag{14}$$

where $\epsilon = \frac{\delta}{(b-c)R}$.

The output of the LIF neuron is assumed to be corrupted by Gaussian white noise $\{\xi_k\}_{k\in\mathbb{Z}}$ with zero mean and standard deviation A_ξ, which models the error associated with the measurement of the spike times $\{t_k\}_{k\in\mathbb{Z}}$.

3.1 An identification method based on an equivalent LF-LIF circuit

As before, the identification of the LF-LIF circuit is carried out in two distinct steps. The first step involves the identification of the LIF neuron, which requires estimating four parameters. By deriving an input-output equivalent LF-LIF circuit, the problem can be simplified, requiring the estimation of only two parameters. The equivalence relation is a consequence of the following lemma.

Lemma 2. *Let $\{t_k\}_{k\in\mathbb{Z}}$ be the spike times sequence generated by neuron $LIF_{\{R,C,\delta,b\}}$ in response to input $v(t)$. Let r be an arbitrary number satisfying $r > \frac{\delta}{R} - b$. Then the following holds true*

$$\int_{t_k}^{t_{k+1}} y(\tau) e^{-\frac{t_{k+1}-\tau}{RC}} d\tau = \delta_b - RC + RC \cdot e^{-\frac{t_{k+1}-t_k}{RC}}, \forall k \in \mathbb{Z},$$

where $\delta_b = \frac{C\delta}{b+r}$ and $y(t) = \frac{v(t)-r}{b+r}$.

Proof. The t-transform of $LIF_{\{R,C,\delta,b\}}$ satisfies (13)

$$\int_{t_k}^{t_{k+1}} v(\tau) e^{-\frac{t_{k+1}-\tau}{RC}} d\tau = C\delta - (b+r)RC\left(1 - e^{-\frac{t_{k+1}-t_k}{RC}}\right) - r \cdot RC\left(1 - e^{-\frac{t_{k+1}-t_k}{RC}}\right)$$

$$\Leftrightarrow \int_{t_k}^{t_{k+1}} (v(\tau) - r)\, d\tau = C\delta - (b+r)(t_{k+1} - t_k). \tag{15}$$

The required result follows after dividing both sides of (15) by $(b+r)$. □

The previous result proves that the neuron $LIF_{\{R,C,\delta,b\}}$ with input $v(t)$ generates the same spike times sequence $\{t_k\}_{k\in\mathbb{Z}}$ as the neuron $LIF_{\{RC,1,\delta_b,1\}}$ with input $y(t)$. In practice, $r = r(u)$ represents, as in Section 2, the steady state output of the filter in response to a step input. As a consequence, it follows that the circuits depicted in Figure 13 and Figure 14 are input-output equivalent.

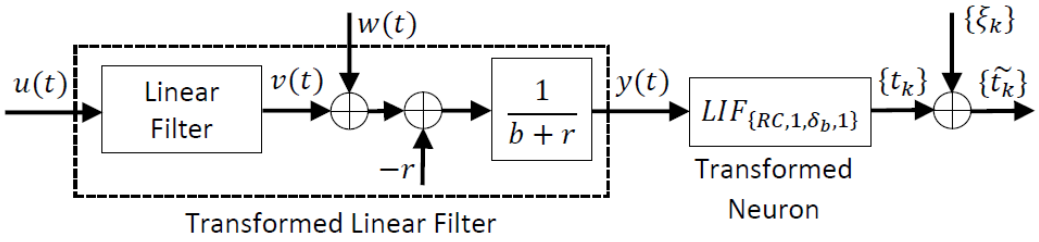

Figure 14: Input-output equivalent LF-LIF circuit.

A method to identify the circuit in Figure 14, which involves first the identification of the spiking neuron followed by the identification of the transformed linear filter, is summarized below.

Step 1. Spiking neuron parameters estimation

The following theorem establishes the basis for the estimation of the spiking neuron parameters. Specifically, it proves that the LIF parameter RC is the unique zero of a function $P(x)$ depending only on the responses of the LF-LIF circuit to a specific set of stimuli. Moreover, the theorem proves that $P(x)$ takes values with opposite signs on each side of RC, which guarantees that the estimator converges to the true value.

Theorem 1. *Let* $\{t_k^n\}_{k=0}^{N_n}, n = 0, 1, 2,$ *be the output spike times sequences generated by*

the LF-LIF circuit in the absence of noise, in response to the following inputs

$$u^n(t) = u^n_\infty \cdot 1_{[0,\infty[}(t), n = 0, 1, 2,$$

where $u^0_\infty = A, u^1_\infty = A - a, u^2_\infty = A + a, A \in \mathbb{R}$ *and* $a \in]0, a_M[$, $a_M = \frac{(b+KA)(RC-\delta_b)}{K \cdot RC}$, *where* K *denotes the filter gain constant. Let* $v^n(t)$ *be the output of the linear filter component of the circuit in response to input* $u^n(t)$, *and let* $y^n(t) = \frac{v^n(t)-KA}{b+KA}$, *for* $n = 0, 1, 2$. *Assuming that the linear filter is BIBO-stable and that the neuron* $LIF_{\{R,C,\delta,b\}}$ *satisfies condition* (14)*, the following hold true*

(a) The limit $\lim_{k\to\infty} \Delta t^n_k = \Delta t^n_\infty$ *exists and is finite,*

where $\Delta t^n_k = t^n_{k+1} - t^n_k, n = 0, 1, 2$;

(b) The spiking neuron parameters satisfy

$$P(x = RC) = 0,$$

$$\delta_b = RC \left[1 - e^{-\frac{-\Delta t^0_\infty}{RC}}\right], \tag{16}$$

where $P(x) = \frac{1-e^{-\frac{\Delta t^n_\infty}{x}}}{1-2e^{-\frac{\Delta t^1_\infty}{x}}+e^{-\frac{\Delta t^0_\infty}{x}}} - \frac{1-e^{-\frac{\Delta t^2_\infty}{x}}}{1-e^{-\frac{\Delta t^0_\infty}{x}}}, \forall x > 0$;

(c) $P(x) = 0$ *has a unique solution;*

(d) $sgn\,(P(x)) = sgn(RC - x), \forall x > 0,$

where $sgn : \mathbb{R} \to \{-1, 0, 1\}$ *is the sign function.*

Proof. See Appendix. □

The assumption $a \in]0, a_M[$, $a_M = \frac{(b+KA)(RC-\delta_b)}{K \cdot RC}$, from Theorem 1 guarantees that the sequences Δt^n_k converge for $n = 0, 1, 2$, as demonstrated in the proof. In practice, if $a \notin]0, a_M[$, it means that the LF-LIF circuit does not generate spikes in response to one

or more of the inputs $u^n(t), n = 0, 1, 2$. In this scenario, the requirement $a \in]0, a_M[$ can be met by adjusting the values A and a.

The parameter RC is obtained by solving $P(x) = 0$ using the bisection method (Courant & Hilbert, 1965). Specifically, the method calculates iteratively sequence $\{x_m\}_{m\in\mathbb{N}}$, $x_m = [x_{m,1} \;\; x_{m,2}]$, where

$$x_{m+1} = \begin{cases} \left[\frac{x_{m,1}+x_{m,2}}{2} \;\; x_{m,2}\right], & P\left(\frac{x_{m,1}+x_{m,2}}{2}\right) > 0, \\ \left[x_{m,1} \;\; \frac{x_{m,1}+x_{m,2}}{2}\right], & P\left(\frac{x_{m,1}+x_{m,2}}{2}\right) < 0, \end{cases} \tag{17}$$

where $x_{0,1}, x_{0,2} \in \mathbb{R}$ denote the initial conditions satisfying $x_{0,1} < x_{0,2}$ and $P(x_{0,1}) \cdot P(x_{0,2}) < 0$. From Theorem 1 (d), it follows that $x_{m,1} < RC < x_{m,2}, \forall m \in \mathbb{N}$, and thus

$$\lim_{m\to\infty} x_{m,i} = RC,$$

for $i = 1, 2$. The parameter δ_b is subsequently determined using equation (16).

In a more realistic scenario assuming the presence of noise and a given finite spike times sequence $\{\widetilde{t}_k^n\}_{k=1}^{N_n}$, the value Δt_∞^n is estimated as

$$\widehat{\Delta t_\infty^n} = \frac{\sum_{k=k_n}^{N_n-1} \widetilde{\Delta t_k^n}}{N_n - k_n}, \tag{18}$$

where $\widetilde{\Delta t_k^n} = \widetilde{t}_{k+1}^n - \widetilde{t}_k^n$, and k_n is the index of spike time $\widetilde{t}_{k_n}^n$, such that, for $\forall k = k_n, \ldots, N_n - 1$, $\left|\widetilde{\Delta t_k^n} - \frac{1}{N_n-k}\sum_{i=k}^{N_n-1} \widetilde{\Delta t_i^n}\right| < \Delta t_{err}^n$, where Δt_{err}^n is a parameter selected by the user.

The neuron parameter RC is computed iteratively using (17). The stop criterion for the iterations is given by $|x_{m,2} - x_{m,1}| < tol_2$, where tol_2 denotes a tolerance value selected by the user.

The estimate of δ_b is given by

$$\widehat{\delta_b} = \widehat{RC}\left[1 - e^{-\frac{\widehat{\Delta t_{\infty}^{0}}}{\widehat{RC}}}\right], \tag{19}$$

where $\widehat{RC}$ denotes the estimation of the neuron parameter RC.

Step 2. Estimation and structure detection of the linear filter

Let $\{t_k\}_{k\in\mathbb{Z}}$ be the output spike times sequence generated by the LF-LIF circuit given the input $u \in PW_{\Omega}, \Omega > 0$. The output $y(t)$ of the transformed linear filter is reconstructed from the spike train $\{t_k\}_{k\in\mathbb{Z}}$ using the method in (Lazar & Pnevmatikakis, 2010), using neuron parameters $\widehat{\delta_b}, \widehat{RC}$ computed in **Step 1**.

In practice functions $u(t)$ and $\hat{y}(t)$ are sampled uniformly with period ε_1. The data is subsequently downsampled with period ε_2, for identification purposes. Let $u[k]$ and $\hat{y}[k]$ be the input and output sequences, sampled with period ε_2, used to identify the linear ARMAX model

$$\begin{aligned}\hat{y}[k] + a_1\hat{y}[k-1] + \cdots + a_{n_y}\hat{y}[k-n_y] = b_1 u[k-1] + \ldots, b_{n_u}u[k-n_u] \\ + e[k] + c_1 e[k-1] + \cdots + c_{n_e}e[k-n_e],\end{aligned}$$

where $e[k]$ is the noise variable and n_u, n_y, n_e are the maximum input, output and noise lags, respectively. The structure of the system is assumed to be unknown, and is identified, as before, using the OFR algorithm (Billings et al., 1989).

3.2 Numerical study

Let $G(s)$ be the transfer function of the linear filter component in the LF-LIF circuit, given by

$$G(s) = \frac{0.8}{0.01s^2 + 0.04s + 1}. \tag{20}$$

The output of the filter is perturbed by additive white Gaussian noise function $w(t)$ with zero mean and standard deviation $\sigma_w = 10^{-2}$. The linear system (20) is connected in cascade with a LIF neuron with parameters $R = 0.02, C = 1, \delta = 0.02$ and $b = 4$. It is assumed that the output of the circuit is measured with no noise, i.e., $A_\xi = 0$.

Step 1. Spiking neuron parameters estimation

The inputs $u^0(t) = 0, u^1(t) = -2 \cdot 1_{[0,\infty[}(t), u^2(t) = 2 \cdot 1_{[0,\infty[}(t), t \in [0, 7\ s]$, sampled with a period $\varepsilon_1 = 10^{-6}\ s$, were used to generate the spike train sequences $\{t_k^n\}_{k=1}^{N_n}, n = 0, 1, 2$, respectively, where $N_0 = 1216, N_1 = 652$, and $N_2 = 1776$. The data was used to determine the spiking neuron parameters following the procedure outlined in Subsection 3.1 above. The parameters $\widehat{\Delta t_\infty^0} = 5.8 \cdot 10^{-3}$, $\widehat{\Delta t_\infty^1} = 10.8 \cdot 10^{-3}$, and $\widehat{\Delta t_\infty^2} = 3.9 \cdot 10^{-3}$ were calculated using (18), where the indices $k_n = 1070$ were calculated for $\Delta t_{err}^n = 8 \cdot 10^{-7}, n = 0, 1, 2$.

The function $P(x)$ and the estimated parameter $\widehat{RC}$ are depicted in Figure 15. Although the function is clearly not monotonic, the bisection method is always convergent due to Theorem 1 (d).

The spiking neuron parameters were estimated as $\widehat{RC} = \frac{x_{40,2} - x_{40,1}}{2} = 0.02003$ (17) and $\widehat{\delta_b} = 5.0008 \cdot 10^{-3}$ (19), where $x_0 = [10^{-3}\ \ 10^4]$ and $tol_2 = 10^{-8}$.

The approximation errors were $e_{RC} = \widehat{RC} - RC = 3.74 \cdot 10^{-5}$ and $e_{\delta_b} = \widehat{\delta_b} - \delta_b = 8.7 \cdot 10^{-7}$, where, in this case, $\delta_b = \delta / b = 5 \cdot 10^{-3}$.

Step 2. Estimation and structure detection of the linear filter

The data used to identify the linear filter was generated by simulating the NF-IIF circuit using an input function $u_{tr}(t)$ with sampling period ε_1 and duration 7 s, whose

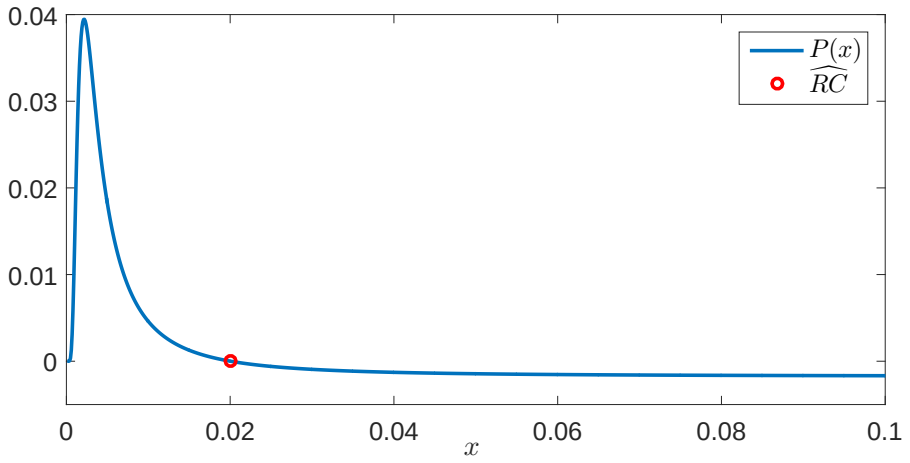

Figure 15: Function $P(x)$ and the estimated value $\widehat{RC}$.

samples are drawn from $N(0,1)$. The input is subsequently low-pass filtered using a Butterworth filter with bandpass corner frequency $30\ rad/s$, stopband corner frequency $50\ rad/s$, maximum attentuation in the passband of $10\ dB$, and minimum attenuation in the stopband of $40\ dB$.

The output of the circuit consisted of a spike time sequence $\{t_k^{tr}\}_{k=1}^{1214}$. To validate the model, a separate input $u_{val}(t)$ and output sequence $\{t_k^{val}\}_{k=1}^{1210}$ were generated using the above procedure.

The data used for estimation was generated by reconstructing the input of the spiking neuron (output of nonlinear filter) from $\{t_k^{tr}\}_{k=1}^{1214}$ and the spiking neuron model identified in step 1, where the sampling period is ε_1. The input/output data was preprocessed to remove the mean, and the first and last 50 samples were discarded, to ensure that the reconstruction distortions due to boundary effects are not affecting the identification procedure. The resulting functions are depicted in Figure 16.

The input and output data used in identification $u_{tr}[k], \hat{y}_{tr}[k]$ was subsequently ob-

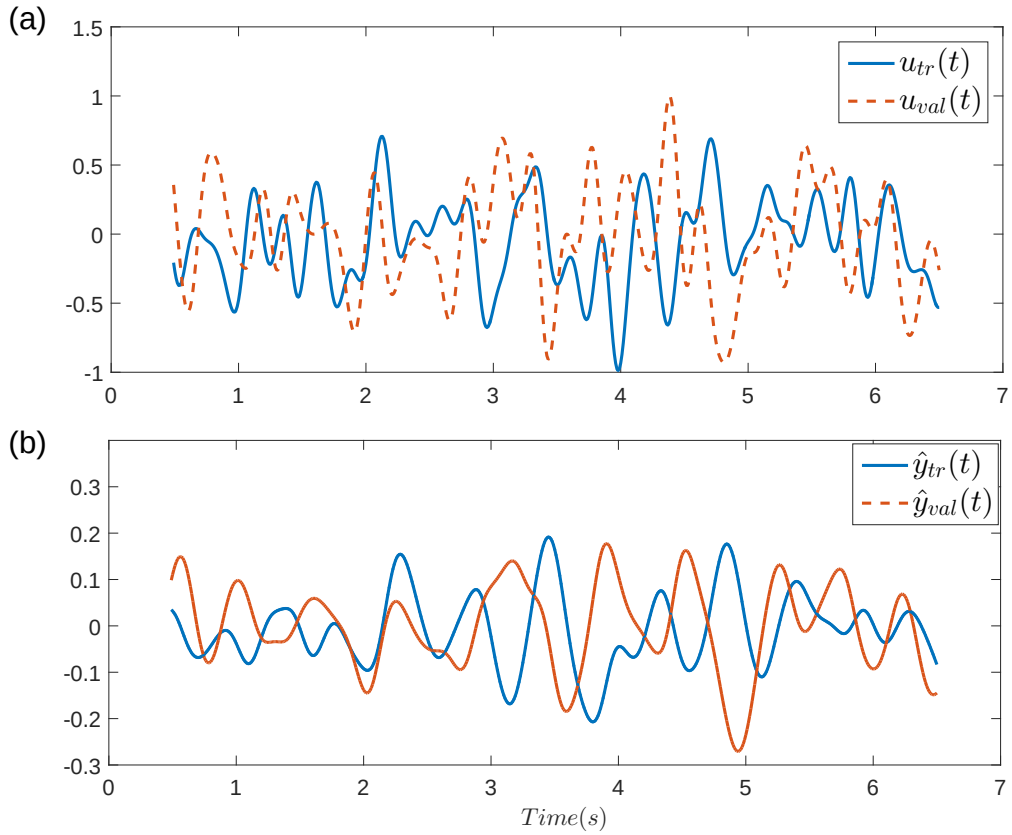

Figure 16: (a) The filter input functions $u_{tr}(t), u_{val}(t)$ and (b) filter outputs $\hat{y}_{tr}(t), \hat{y}_{val}(t)$ used to estimate and validate the ARMAX model.

tained by downsampling the original data with $\varepsilon_2 = 10^{-2}$. The maximum number of lags used in identification are $n_u = n_y = 10, n_e = 0$. The model terms selection and parameter estimation of an ARMAX model was performed using input $u_{tr}[k]$ and output $\hat{y}_{tr}[k]$ with the stop criterion $NMSE < 10^{-3}$. The final set of regressors and the corresponding estimated parameters are presented in Table 2.

The model predicted output $\hat{\hat{y}}_{val}[k]$ corresponding to the validation input $u_{val}[k]$ is

Table 2: The model terms selection and parameter estimation results.

Index s	Model term $p_s(X[k])$	Parameter θ_s	ERR (%)
1	$\hat{y}[k-1]$	1.95	98.98
2	$\hat{y}[k-2]$	-0.96	1.009
3	$u[k-1]$	$2 \cdot 10^{-3}$	$4.92 \cdot 10^{-3}$

shown in Figure 17a. The model prediction error $e_{val}[k]$ is shown in Figure 17b. The NMSE for training and validation are $3.004 \cdot 10^{-5}$ and $3.23 \cdot 10^{-5}$, respectively.

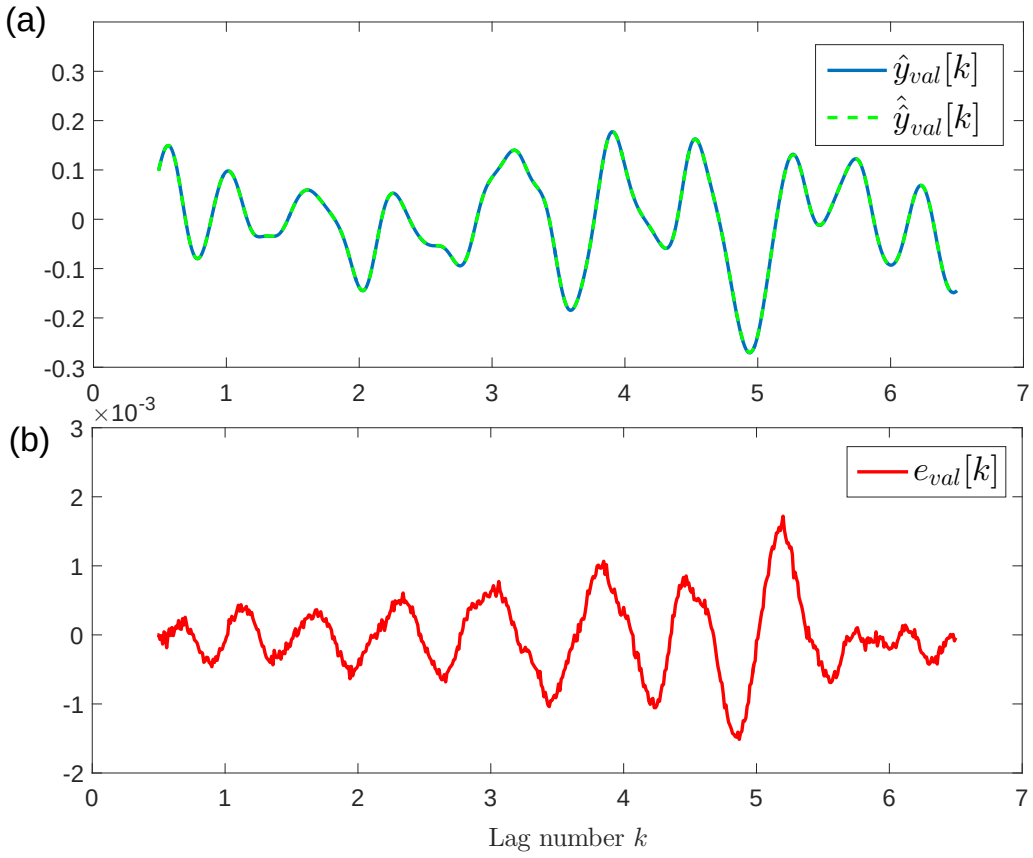

Figure 17: (a) Validation $\hat{y}_{val}[k]$ and model predicted output $\hat{\hat{y}}_{val}[k]$. (b) Prediction error $e_{val}[k]$.

To validate the ARMAX model, the linear frequency response function of the iden-

tified model $\hat{G}(j\omega)$ was compared to the one of the original system. The magnitude frequency response function of the original system $G(j\omega)$ is shown in Figure 18a. The magnitude error function $E_1(\omega)$ is computed (7), and depicted in Figure 18b.

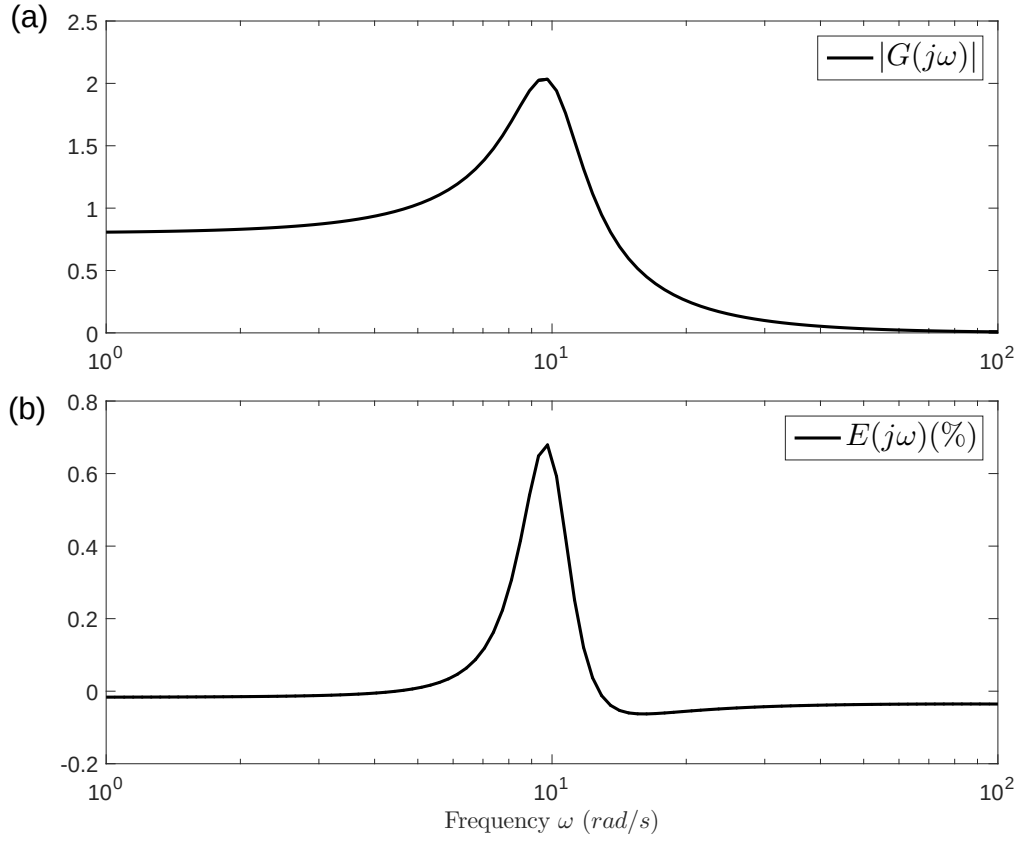

Figure 18: (a) The magnitude frequency response function $G(j\omega)$ associated with the linear system; (b) The magnitude error function $E_1(j\omega)$.

Moreover, the identified circuit was validated in terms of the spiking output, by simulating its response to inputs $u_{tr}(t)$ and $u_{val}(t)$. In order to simulate the linear filter, a new set of inputs $u_{tr}^{\varepsilon_2,i}[k], u_{val}^{\varepsilon_2,i}[k], i = 1, \ldots, 10^4$, were generated as in Example 1 (9).

The outputs of the linear filter $\hat{\hat{y}}_{tr}(t), \hat{\hat{y}}_{val}(t)$, sampled with ε_1, are subsequently computed from the responses of the filter with inputs $u_{tr}^{\varepsilon_2,i}[k]$, $u_{val}^{\varepsilon_2,i}[k]$, respectively (10). The generated output spike times were validated against the original spike time values using the coincidence factor with precision $\Delta = 0.0015\ s$. The values of the coincidence factor for training and validation were $\Gamma_{tr} = 1$ and $\Gamma_{val} = 1$, respectively.

This methodology can be used for the identification of a linear filter connected via multiplicative coupling to a Hodgkin-Huxley model (LF-HH). This results in a very large resistance value for the LIF, which essentially turns it into a IIF model. Moreover, the neuron identified in Example 4, Section 2, was shown to have significantly nonlinear subthreshold dynamics, thus making it unsuitable for this methodology.

4 Conclusions

The paper introduced two novel identification methodologies for circuits consisting of filters in cascade with spiking neurons. The first approach concerns circuits consisting of nonlinear filters in cascade with IIF neurons. The second approach is suitable for circuits comprising linear filters in series with LIF neurons. Compared to the previous approaches, the methods do not require *a priori* knowledge of the spiking neuron parameters or the filter structure, and do not assume that the input of the neuron (i.e. the output of the filter) is available for measurement. Both approaches are based on an equivalent representation of the circuit, which decreases the number of tunable parameters. The identification procedure involves two steps: the estimation of the spiking neuron and the identification of the filter.

Numerical simulations are used to demonstrate the applicability and performance of

both proposed methodologies in the presence of additive noise applied to the output of the filter as well as to the measured spike time sequence.

In the case of the NF-IIF circuit, the proposed identification method addresses the well known limitations of the Volterra-based identification approaches. In particular, the proposed approach can be used to identify NF-IIF circuits where the nonlinear filter is not memoryless, and can even be chaotic. It is also shown that the identification approach can be used to infer an equivalent NF-IIF model of circuits incorporating a Hodgkin-Huxley model. The proposed identification method was also demonstrated using a real experimental data set from the Allen Cell Type Database. It is shown that the proposed approach can be used to identify neuron models that reproduce robustly the experimental data.

In the case of sensory circuit models incorporating the LIF spiking neuron model, identifying the parameters of the neuron is performed under the assumption that the filter is linear. This allows estimating the two parameters of the equivalent LIF neuron from the output spike time sequences corresponding to three step inputs. This method trades off the generality of a nonlinear filter for a more general model of the spiking neuron.

In essence, the proposed approaches allow identifying computational models that can characterize the neural computations performed by early sensory circuits incorporating graded-potential as well as spiking neurons. These models can be connected to models of downstream neural circuits that are identified subsequently based on recordings made in the downstream spiking neurons. This provides a route to constructing more complex models of early sensory processing.

Appendix A

Identification results

Table 3: The model terms selection and parameter estimation results for the nonlinear system in Subsection 2.2, Example 1. The terms are given in the descending order of their error reduction ratios (ERRs), which show the percentage contribution of the term to the model output.

Index s	Model term $p_s(X[k])$	Parameter θ_s	ERR $(\%)$
1	$\hat{y}[k-1]$	1.12	98.01
2	$\hat{y}[k-2]$	-0.56	1.96
3	$u[k-1]$	$1.1 \cdot 10^{-3}$	$8.6 \cdot 10^{-3}$
4	$\hat{y}[k-4]$	-0.91	$3.2 \cdot 10^{-3}$
5	$\hat{y}[k-3]\hat{y}[k-1]$	-0.14	$2.2 \cdot 10^{-3}$
6	$u[k-3]$	$4.4 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
7	$\hat{y}[k-3]$	1.12	$8 \cdot 10^{-4}$
8	$\hat{y}[k-5]$	0.66	$5 \cdot 10^{-4}$
9	$\hat{y}[k-6]$	-0.53	$2 \cdot 10^{-4}$
10	const.	$1 \cdot 10^{-4}$	$2 \cdot 10^{-4}$

Table 4: The model terms selection and parameter estimation results for the nonlinear system in Subsection 2.2, Example 2.

Index s	Model term $p_s(X[k])$	Parameter θ_s
1	$\hat{y}[k-1]$	3.55
2	$\hat{y}[k-2]$	-4.9
3	$\hat{y}[k-3]$	2.91
4	$\hat{y}[k-4]$	-0.12
5	$\hat{y}[k-6]$	0.29
6	$\hat{y}[k-9](\hat{y}[k-1])^2$	-0.69
7	$u[k-10](\hat{y}[k-1])^2$	$9.55 \cdot 10^{-3}$
8	$u[k-10]\hat{y}[k-1]\hat{y}[k-2]$	$-2.3 \cdot 10^{-2}$
9	$u[k-10]\hat{y}[k-1]\hat{y}[k-3]$	$1.39 \cdot 10^{-2}$
10	$u[k-7]\hat{y}[k-10]\hat{y}[k-1]$	$-1.88 \cdot 10^{-4}$
11	$\hat{y}[k-10]\hat{y}[k-1]\hat{y}[k-2]$	1.7
12	$u[k-1]$	$-9 \cdot 10^{-5}$
13	$(\hat{y}[k-1])^3$	-6.28
14	$\hat{y}[k-10]\hat{y}[k-4]\hat{y}[k-2]$	1.15
15	$u[k-10]\hat{y}[k-4]\hat{y}[k-1]$	$-3.05 \cdot 10^{-4}$
16	$\hat{y}[k-10]\,(\hat{y}[k-2])^2$	-2.41
17	$\hat{y}[k-2](\hat{y}[k-1])^2$	13.96
18	$\hat{y}[k-3](\hat{y}[k-1])^2$	-8.39
19	$u[k-2]$	$1.68 \cdot 10^{-4}$
20	$\hat{y}[k-8](\hat{y}[k-1])^2$	1.04
21	$\hat{y}[k-10]$	$-1.72 \cdot 10^{-2}$
22	$\hat{y}[k-9]\hat{y}[k-8]\hat{y}[k-3]$	-0.37
23	$\hat{y}[k-5]$	-0.72

Table 5: The model terms selection and parameter estimation results for the nonlinear system in Subsection 2.2, Example 3.

Index s	Model term $p_s(X[k])$	Parameter θ_s
1	$\hat{y}[k-1]$	6.97
2	$\hat{y}[k-2]$	-23.68
3	$\hat{y}[k-3]$	51.3
4	$\hat{y}[k-4]$	-78.07
5	$\hat{y}[k-5]$	86.9
6	$\hat{y}[k-6]$	-71.5
7	$\hat{y}[k-7]$	42.89
8	$\hat{y}[k-8]$	-17.93
9	$\hat{y}[k-9]$	4.71
10	$\hat{y}[k-10]$	-0.58
11	$u[k-5]$	$-3.53 \cdot 10^{-4}$
12	$u[k-1]$	$3.68 \cdot 10^{-4}$
13	$u[k-2]$	$-5.84 \cdot 10^{-4}$
14	$\hat{y}[k-10]\hat{y}[k-1]$	$4.32 \cdot 10^{-4}$
15	$u[k-6]\hat{y}[k-1]$	$-1.84 \cdot 10^{-5}$
16	$u[k-10]$	$1.24 \cdot 10^{-5}$
17	$u[k-4]$	$6.08 \cdot 10^{-4}$
18	$(u[k-1])^2$	$-1.36 \cdot 10^{-3}$
19	const.	$2.63 \cdot 10^{-6}$

Table 6: The model terms selection and parameter estimation results for the nonlinear system in Subsection 2.2, Example 4.

Index s	Model term $p_s(X[k])$	Parameter θ_s	ERR $(\%)$
1	$u[k-2]$	0.0054	65.4
2	$(u[k-2])^2$	$-8.44 \cdot 10^{-6}$	5.43
3	$u[k-3]$	-0.0016	4.41
4	$u[k-1]$	0.0995	2.53

Appendix B

Proofs of theorems

The following auxiliary lemma is used in the proof of Theorem 1 (c) and (d).

Lemma 3. *Let* $\Lambda_z :]1, +\infty[\rightarrow]1, +\infty[, \Lambda_z(s) = \frac{1}{1-\left(1-\frac{1}{s}\right)^z}, z \in]0, +\infty[$.

Then $\Lambda_z(s)$ *is strictly concave for* $z < 1$*, and strictly convex for* $z > 1$.

Proof. The following holds true.

$$\Lambda_z'(s) = \frac{z}{s^2\left(\left(\frac{s}{s-1}\right)^{\frac{z-1}{2}} - \left(\frac{s-1}{s}\right)^{\frac{z+1}{2}}\right)^2} = z\frac{((s-1)s)^{z-1}}{\left(s^z - (s-1)^z\right)^2},$$

and

$$\Lambda_z''(s) = \frac{((s-1)s)^{z-2}\, h(s)}{(s^z - (s-1)^z)^3},$$

where

$$h(s) = (z-1)(2s-1)(s^z - (s-1)^z) - 2zs(s-1)(s^{z-1} - (s-1)^{z-1}).$$

It is easy to see that $\operatorname{sgn}(\Lambda_z''(s)) = \operatorname{sgn}(h(s)), \forall s \in]1, +\infty[$. After simple calculations, it follows that

$$h(s) = (s-1)^z(2s+z-1) - s^z(2s-z-1)$$

$$= s^z(2s+z-1)\left(\left(\frac{s-1}{s}\right)^z - \frac{2s-z-1}{2s+z-1}\right).$$

To assess the sign of $h(s)$, the following function is evaluated

$$h\left(\lambda(p)\right) = \left(\frac{1}{1-p}\right)^z \left(\frac{2}{1-p}+z-1\right)\left(p^z - \frac{p(1+z)+(1-z)}{p(1-z)+(1+z)}\right), \tag{21}$$

where $\lambda :]0,1[\rightarrow]1,+\infty[$, $\lambda(p) \triangleq \frac{1}{1-p}, \forall p \in]0,1[$. The following holds

$$\left(\frac{1}{1-p}\right)^z \left(\frac{2}{1-p}+z-1\right) > 0, \forall p \in]0,1[.$$

Case I. $z < 1$.

In this case $p(1+z)+(1-z) > 0, \forall p, z \in]0,1[$. It follows that $\text{sgn}\left(h(\lambda(p))\right) = \text{sgn}(\theta(p)), \forall p \in]0,1[$, where $\theta :]0,1[\rightarrow \mathbb{R}$,

$$\theta(p) \triangleq z \cdot \ln(p) - \ln\left(\frac{p(1+z)+(1-z)}{p(1-z)+(1+z)}\right),$$

such that $p(1-z)+(1+z) > 0, \forall p \in]0,1[$. Furthermore,

$$\begin{aligned}\theta'(p) &= \frac{z}{p} - \frac{4z}{\left(p(1-z)+(1+z)\right)\left(p(1+z)+(1-z)\right)} \\ &= \frac{z(1-z^2)(p-1)^2}{p\left(p(1-z)+(1+z)\right)\left(p(1+z)+(1-z)\right)}.\end{aligned}$$

Then $\theta'(p) > 0, \forall p \in]0,1[$ and $\lim_{p\rightarrow 1}\theta(p) = 0$. It follows that $\theta(p) < 0, h(\lambda(p)) < 0, \forall p \in]0,1[, h(s) < 0, \Lambda_z''(s) < 0, \forall s \in]1,+\infty[$, and thus the lemma holds true.

Case II. $z > 1$. The following holds.

$$p(1+z)+(1-z) \leq 0, p \in]0, p_0]$$

$$p(1+z)+(1-z) > 0, p \in]p_0, 1[\,.$$

where $p_0 = \frac{z-1}{z+1}$.

For $p \in]0, p_0]$ it follows that $h(\lambda(p)) > 0, \forall p \in]0, p_0]$ (21), and thus $h(s) > 0, \Lambda_z''(s) > 0, \forall s \in]1, \frac{1}{1-p_0}[$.

For $p \in]p_0, 1[$, $\theta(p)$ and $\theta'(p)$ are calculated as in **Case I.** Then $\theta'(p) < 0, \lim_{p\to 1} \theta(p) = 0$, $\theta(p) > 0, \forall p \in]p_0, 1[$, and the result follows. □

Proof of Theorem 1.

(a) Because the linear system is BIBO-stable, it follows that $\exists \lim_{t\to\infty} v^n(t) = v_\infty^n = K \cdot u_\infty^n, n = 0, 1, 2$, and thus

$$y_\infty^n = \lim_{t\to\infty} y^n(t) = \lim_{t\to\infty} \frac{v^n(t) - KA}{b + KA} = \frac{Ku_\infty^n - KA}{b + KA} = K_b \left(u_\infty^n - A\right), \tag{22}$$

for $n = 0, 1, 2$, where $K_b = \frac{K}{b+KA}$.

Let Δt_∞^n denote one of the limits of the sequence Δt_k^n, for every $n = 0, 1, 2$. This limit will be proven to be unique in the following. The following holds from Lemma 2 with neuron inputs $v^n(t)$, neuron output sequences $\{t_k^n\}_{k=1}^{N_n}, n = 0, 1, 2$, and $r = KA$

$$\int_{t_k^n}^{t_{k+1}^n} \left(y^n(\tau) + 1\right) e^{-\frac{t_{k+1}^n - \tau}{RC}} d\tau = \delta_b,$$

for $\forall k \in \mathbb{Z}$, where $\delta_b = \frac{C\delta}{b+KA}$. Therefore, it follows that

$$\begin{aligned}
\delta_b &= \lim_{k\to\infty} \int_{t_k^n}^{t_{k+1}^n} \left(y^n(\tau) + 1\right) e^{-\frac{t_{k+1}^n - \tau}{RC}} d\tau \\
&\underset{\tau=(t_{k+1}^n-\zeta)}{=} \lim_{k\to\infty} \int_0^{\Delta t_k^n} \left(y^n(t_{k+1}^n - \zeta) + 1\right) e^{-\frac{\zeta}{RC}} d\zeta \\
&= \int_0^{\Delta t_\infty^n} \left(y_\infty^n + 1\right) \cdot e^{-\frac{\zeta}{RC}} d\zeta \\
&= \left(y_\infty^n + 1\right) RC \left[1 - e^{-\frac{\Delta t_\infty^n}{RC}}\right] \\
&\overset{(22)}{=} \left(K_b \left(u_\infty^n - A\right) + 1\right) RC \left[1 - e^{-\frac{\Delta t_\infty^n}{RC}}\right],
\end{aligned} \tag{23}$$

for $n = 0, 1, 2$, or, equivalently,

$$\Delta t_\infty^n = -RC \cdot \ln\left(1 - \frac{\delta_b}{RC(K_b \left(u_\infty^n - A\right) + 1)}\right). \tag{24}$$

It follows that Δt_∞^n exists and satisfies $\Delta t_\infty^n \in]0, \infty[$ if and only if

$$\frac{\delta_b}{RC(K_b\,(u_\infty^n - A) + 1)} < 1 \Leftrightarrow u_\infty^n - A > \frac{\delta_b - RC}{K_b \cdot RC}, \tag{25}$$

for $n = 0, 1, 2$, or, equivalently, $a < a_M = \frac{RC - \delta_b}{K_b \cdot RC}$.

(b) The following holds true (23)

$$\delta_b = RC \left[1 - e^{-\frac{\Delta t_\infty^0}{RC}}\right]. \tag{26}$$

The expression of δ_b in (26) is substituted in (23) for $n = 1$

$$K_b = \frac{1}{a}\left[1 - \frac{1 - e^{-\frac{\Delta t_\infty^0}{RC}}}{1 - e^{-\frac{\Delta t_\infty^1}{RC}}}\right]. \tag{27}$$

Finally, the result follows after the substitution of (27) and (26) in (23), for $n = 2$.

(c) The values $\Delta t_\infty^n, n = 0, 1, 2$, satisfy (24)

$$\Delta t_\infty^n = -RC \cdot \ln(d_n), \forall n = 0, 1, 2,$$

where $d_n \triangleq 1 - \frac{\delta_b}{RC(K_b(u_\infty^n - A)+1)}, n = 0, 1, 2$. It follows that $0 < d_2 < d_0 < d_1 < 1$ (25). Function $P(x)$ satisfies

$$P(x) = \frac{1 - d_1^{\frac{RC}{x}}}{1 - 2d_1^{\frac{RC}{x}} + d_0^{\frac{RC}{x}}} - \frac{1 - d_2^{\frac{RC}{x}}}{1 - d_0^{\frac{RC}{x}}}. \tag{28}$$

It can be easily verified that $x = RC$ is a solution to $P(x) = 0$, by substituting the expressions of $\{d_n\}_{n=0,1,2}$ in (28).

Let $z > 0, z \triangleq \frac{RC}{x}$. Then the following holds

$$\begin{aligned} P(x) = 0 &\Leftrightarrow \frac{1 - 2d_1^z + d_0^z}{1 - d_1^z} = \frac{1 - d_0^z}{1 - d_2^z} \\ &\Leftrightarrow \frac{(2 - 2d_1^z) - (1 - d_0^z)}{1 - d_1^z} = \frac{1 - d_0^z}{1 - d_2^z} \\ &\Leftrightarrow \frac{2}{1 - d_0^z} = \frac{1}{1 - d_1^z} - \frac{1}{1 - d_2^z}. \end{aligned} \tag{29}$$

Let $\lambda :]0,1[\rightarrow]1,+\infty[$ be a strictly increasing and continuous function with expression $\lambda(p) \triangleq \frac{1}{1-p}, \forall p \in]0,1[$. Then λ is a one-to-one and onto function, and thus it has an inverse $\lambda^{-1} :]1,+\infty[\rightarrow]0,1[, \lambda^{-1}(s) = 1 - \frac{1}{s}$. Equation (29) is satisfied for $z = 1$, such that

$$2\lambda(d_0) = \lambda(d_1) + \lambda(d_2).$$

Let $s_n = \lambda(d_n), n = 0,1,2$, and let $\Lambda_z :]1,+\infty[\rightarrow]1,+\infty[, \Lambda_z(s) = \lambda(\lambda^{-1}(s)^z)$. Then the following holds true (29)

$$2s_0 = s_1 + s_2,$$

$$2\Lambda_z(s_0) = \Lambda_z(s_1) + \Lambda_z(s_2).$$

The function Λ_z is strictly convex for $z > 1$, due to Lemma 3, and thus it is also strictly midpoint convex, i.e.,

$$2\Lambda_z(s_0) < \Lambda_z(s_1) + \Lambda_z(s_2), z > 1. \tag{30}$$

Similarly, Lemma 3 proves that Λ_z is strictly concave for $z < 1$, and thus

$$2\Lambda_z(s_0) > \Lambda_z(s_1) + \Lambda_z(s_2), z < 1. \tag{31}$$

Therefore there is a unique solution $z \in]0,+\infty[$ to equation (29).

(d) The following holds, which concludes the proof

$$\begin{aligned}
\text{sgn}\left(P(x)\right) &= -\text{sgn}\left(\frac{1 - 2d_1^{\frac{RC}{x}} + d_0^{\frac{RC}{x}}}{1 - d_1^{\frac{RC}{x}}} - \frac{1 - d_0^{\frac{RC}{x}}}{1 - d_2^{\frac{RC}{x}}}\right) \\
&= -\text{sgn}\left(2\Lambda_{\frac{RC}{x}}(s_0) - \Lambda_{\frac{RC}{x}}(s_1) + \Lambda_{\frac{RC}{x}}(s_2)\right) \\
&\overset{(30),(31)}{=} -\text{sgn}(x - RC), \forall x \in]0,+\infty[.
\end{aligned}$$

□

Acknowledgments

DF and DC gratefully acknowledge that this work was supported by BBSRC under grant BB/M025527/1.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. In *Annals of the Institute of Statistical Mathematics*, *21* (1): 243 – 247.

Allen Institute for Brain Science (2015). Allen Cell Types Database. Available from: http://celltypes.brain-map.org/.

Allen Institute for Brain Science (2016). Single neuron electrophysiology recordings. Available from: http://celltypes.brain-map.org/mouse/experiment/electrophysiology/320466871.

Billings, S. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains.* John Wiley & Sons.

Billings, S. A., & Aguirre, L. A. (1993). Validating identified nonlinear models with chaotic dynamics. *International journal of Bifurcation and Chaos*, *4* (1), 109 – 125.

Billings, S. A., & Aguirre, L. A. (1995). Effects of the sampling time on the dynamics and identification of nonlinear models. *International journal of Bifurcation and Chaos*, *5* (6), 1541 – 1556.

Billings, S., Korenberg, M., & Chen, S. (1989). Identification of non-linear output-

affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, *19* (8), 1559 – 1568.

Billings, S. A., & Chen, S. (1989). Extended model set, global data and threshold model identification of severely non-linear systems. *International Journal of Control*, *50* (5), 1897 – 1923.

Billings, S., Chen, S., & Korenberg, M. (1989). Identification of mimo non-linear systems using a forward-regression orthogonal estimator. *International Journal of Control*, *49* (6), 2157 – 2189.

Brette, R. & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of neurophysiology*, *94*(5), 3637–3642.

Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, *95* (1), 1–19.

Chen, S., & Billings, S. (1989). Representations of nonlinear systems: the NARMAX model. *International Journal of Control*, *49* (3), 1013 – 1032.

Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, *50* (5), 1873 – 1896.

Clopath, C., Jolivet, R., Rauch, A., Lüscher, H. R., & Gerstner, W. (2007). Predicting neuronal activity with simple models of the threshold type: Adaptive exponential

integrate-and-fire model with two compartments. *Neurocomputing*, *70*(10), 1668-1673.

Coca, D., Zheng, Y., Mayhew, J. E. W. & Billings, S. A. (2000). Nonlinear system identification and analysis of complex dynamical behavior in reflected light measurements of vasomotion. In *International Journal of Bifurcation and Chaos*, *10* (2), 461 – 476.

Cochran D., & Clark J. J. (1965). *Methods of mathematical physics (Vol. 1).* CUP Archive.

Friederich, U., Billings, S., Hardie, R. C., Juusola, M. & Coca, D. (2016). Fly photoreceptors encode phase congruency. In *PLOS ONE*, *11* (6).

Gabbiani, F. & Cox, S. J. (1998). *Mathematics for Neuroscientists.* Academic Press.

George, D. A. (1959). *Continuous nonlineary systems* (Report No. TR-355) Massachusetts Inst of Tech Cambridge Research Lab of Electronics, 1959.

Gerstner, W, Kistler, WM, Naud, R, & Paninski, L (2014). *Neuronal Dynamics.* Cambridge University Press, Cambridge, 2014.

Gerstner, W. & Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity.* Cambridge University Press, Cambridge, 2002.

van Hemmen, J. L. & Ritz, R. (1995). Neural coding: A theoretical vista of mechanisms, techniques, and applications. In *Analysis of Dynamical and Cognitive Systems*: 75-119, Springer Berlin Heidelberg.

Herz, A. V. M., Gollisch, T., Machens, C. K. & Jaeger, D.(2006). Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction. *Science*, *314*: 80–85.

Hodgkin, A. L. & Huxley, A. F.(1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, *117* (4), 500.

Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological cybernetics*, *55* (2), 135-144.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience.* MIT press.

Jolivet, R., Rauch, A., Lüscher, H. R., & Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. In *Journal of computational neuroscience*, *21* (1), 35–49.

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: a model for the responses of visual neurons. In *Neuron*, *30* (3), 803–817.

Kistler, W. M., Gerstner, W. & van Hemmen, J. L. (1997). Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. In *Neural Computation*, *9* (5), 1015-1045.

Koch, C. & Segev, I. (1998). *Methods in neuronal modeling: from ions to networks.* MIT press.

Lapicque, L. (1907). Recherches Quantitatives sur l'Excitation Électrique des Nerfs Traiteé comme une Polarization. *J. Physiol. Pathol. Gen.*, *9*, 620 – 635.

Lau, B, Stanley, GB, & Dan, Y (2002). Computational subunits of visual cortical neurons revealed by artificial neural networks. In *Proceedings of the National Academy of Sciences*, *99* (13): 8974-8979.

Lazar, A. A. (2010). Population encoding with hodgkin?huxley neurons. *IEEE Transactions on Information Theory*, *56*(2), 821–837.

Lazar, A. A. (2005). Multichannel time encoding with integrate-and-fire neurons. *Neurocomputing*, *65*, 401 – 407.

Lazar, A. A., & Pnevmatikakis, E. A. (2008). Faithful Representation of Stimuli with a Population of Integrate-and-Fire Neurons. *Neural Computation*, *20* (11), 2715–2744.

Lazar A. A., & Tóth T. L. (2003). Time Encoding and Perfect Recovery of Bandlimited Signals. In *1990 Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, *6*, 709–712.

Lazar, A. A., & Slutskiy, Y. B.(2010). Identifying dendritic processing. In *Advances in Neural Information Processing Systems*, *23*: 1261–1269, 2010.

Lazar, A. A., & Pnevmatikakis, E. A. (2010). Consistent recovery of sensory stimuli encoded with MIMO neural circuits. *Computational intelligence and neuroscience*, *2010* (2).

Lazar, A. A., & Pnevmatikakis, E. A. (2011). Video time encoding machines. *IEEE Transactions on Neural Networks*, *22* (3): 461–473, 2011.

Lazar, A. A., Pnevmatikakis, E. A., & Zhou , Y. (2011). Encoding Natural Scenes with Neural Circuits with Random Thresholds. *Vision Research*, *50* (22): 2200–2212, 2010.

Lazar, A. A., & Slutskiy, Y. B. (2015). Spiking neural circuits with dendritic stimulus processors. *Journal of Computational Neuroscience*, *38* (1): 1–24, 2015.

Lazar, A. A., & Slutskiy, Y. B. (2014). Functional identification of spike-processing neural circuits. *Neural computation*, *26* (2): 264–305, 2014.

Leontaritis, IJ & Billings, SA (1981). Identification of non-linear systems using parameter estimation techniques. In *Proceedings of IEEE Conference of Control and Applications*, 183 – 190.

Lesica, NA & Stanley, GB (2004). Encoding of Natural Scene Movies by Tonic and Burst Spikes in the Lateral Geniculate Nucleus. In *The Journal of Neuroscience*, *24* (47): 10731–10740.

Li, L., M., & Billings, S., A.(2011). Estimation of generalized frequency response functions for quadratically and cubically nonlinear systems. *Journal of Sound and Vibration*, *330* (3), 461 – 470.

Millard, DC, Wang, Q, Gollnick, CA & Stanley, GB (2013). System identification of the nonlinear dynamics in the thalamocortical circuit in response to patterned thalamic microstimulation in-vivo. In *The Journal of Neural Engineering*, *10* (6).

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. In *Network: Computation in Neural Systems*, *15* (4), 243–262.

Paninski, L., Pillow, J. W., & Simoncelli, E. P. (2004). Maximum Likelihood Estimation of a Stochastic Integrate-and-Fire Neural Encoding Model. In *Neural Computation*, *16*(12): 2533?-2561.

Pillow, J. (2007). Likelihood-based approaches to modeling the neural code. *Bayesian brain: Probabilistic approaches to neural coding*, 53–70.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. In *The Journal of Neuroscience*, *25*(47): 11003 – 11013.

Simoncelli, E. P., Paninski, L., Pillow, J. & Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. *The cognitive neurosciences*, *3*, 327–338.

Smith, C. (2008). *Biology of sensory systems.* John Wiley & Sons.

Song, D., Robinson, B., Hampson, R., Marmarelis, V., Deadwyler, S., & Berger, T. (2016). Sparse Large-Scale Nonlinear Dynamical Modeling of Human Hippocampus for Memory Prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.*

Smith, C. (2008). *Biology of sensory systems.* John Wiley & Sons.

Swain, A. K. & Billings, S. A.(1998) Weighted complex orthogonal estimator for identifying linear and non-linear continuous time models from generalised frequency response functions. *Mechanical Systems and Signal Processing*, *12*(2): 269 – 292.

Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology.*, Cambridge: Cambridge University Press.

Ueda, Y. (1985). Random phenomena resulting from non-linearity in the system described by Duffing's equation. *International journal of non-linear mechanics*, *20* (5), 481 – 491.

Wei, H., L., Zheng, Y., Pan, Y., Coca, D., Li, L., M., Mayhew, J., E., & Billings, S. (2009). Model estimation of cerebral hemodynamics between blood flow and volume changes: a data-based modeling approach. *IEEE Transactions on Biomedical Engineering*, *56* (6), 1606 – 1616.

Wu, M. C. K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory neurons by systems identification. *Annu.Rev.Neurosci.*, *29*: 477–505, 2006.