

Bio-Inspired Approach to Modelling Retinal Ganglion Cells using System Identification Techniques

P. Vance, G. P. Das, D. Kerr, S. A. Coleman and T. M. McGinnity

Abstract— The processing capabilities of the biological vision system are still vastly superior to artificial vision, which has been an active area of research for over half a century. Current artificial vision techniques, motivated by this robust performance, integrate many insights from biology yet they remain far-off the capabilities of animals and humans in terms of speed, power and performance. With respect to modelling the retina, this is due to an insufficient understanding of the complex interactions between the cells and their organisation within the system. The core components within this system are the retinal ganglion cells as they convey the accumulated data of real world images as action potentials onto the visual cortex via the optic nerve. Computational models that approximate the processing that occurs within these visual neurons can be derived by quantitatively fitting particular sets of physiological data using an input-output analysis where the input is a known and its output is recorded. Techniques capable of mapping this input-output response involve computational combinations of linear and nonlinear models that are generally complex and lack any relevance to the underlying biophysics. In this work we illustrate how system identification techniques, which take inspiration from biological systems, can accurately model ganglion cell behaviour, and are a viable alternative to traditional linear-nonlinear approaches.

Index Terms—Retinal ganglion cells, computational modelling, biological vision, receptive field, artificial stimuli.

I. INTRODUCTION

Mimicking biological vision systems has been a consistent challenge in the visual research field for many years. Vision begins with light that is projected to the back of the eye onto the retina which is an extension of the brain approximately 0.3-0.4mm thick and covers an area of approximately 520mm² [1]. Around 125 million rod and cone photoreceptors transform visible light into neural signals [2]. This is in comparison to 1 million ganglion cells which receive the signal information, having been filtered through intermediate layers consisting of horizontal, bipolar and amacrine cells. There are around 15-20 distinct types of retinal ganglion cells (RGC) which transform the signal information into what are known as action potentials (spikes) and transmit the information via synaptic connections to the visual cortex

Manuscript submitted March 09 2016. This work is supported by the European Union Seventh Framework Programme (FP7-ICT-2011.9.11) under grant number [600954] (“VISUALISE”)

All authors are with the Intelligent Systems Research Centre, University of Ulster, Magee Campus, N. Ireland. (e-mail: {p.vance, g.das, d.kerr, sa.coleman, tm.mcginny}@ulster.ac.uk). T. M. McGinnity is also with Nottingham Trent University, UK.

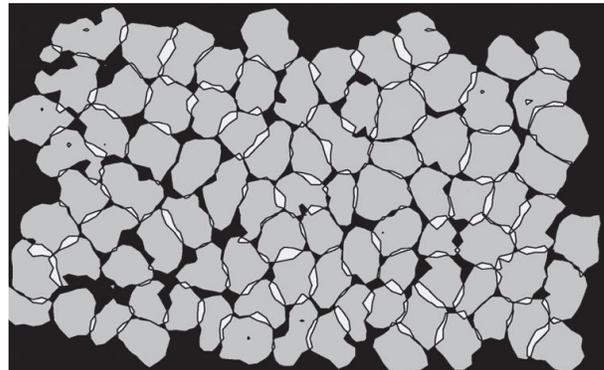


Figure 1: Recorded receptive fields from 'parasol' ganglion cells in macaque retina [1]

for higher processing. Previously, the retina was thought of as a simple spatiotemporal filter, with the real processing beginning in the visual cortex. However this view has been substantially revised in recent times [3].

There is very little feedback from the brain to the retina thus it is an ideal biological system to derive computational models of a stimulus-response relationship, as the inputs can be precisely controlled whilst the output can be extracellularly recorded from RGCs through the use of a multi-electrode array [4]. Each RGC pools signals from multiple photoreceptors via a networked infrastructure of the various cell types. Collectively, the spatial area of photoreceptors which contribute to a RGC eliciting a response is known as the receptive field (RF), which can also be referred to as the region of the sensory space in which visual stimulus triggers a neuron to fire. The general shape of this spatial area is commonly approximated to be either a circular [5] or an elliptical region that is often defined with a 2D Gaussian spatial profile [6], [7]. In reality, however, the actual shape of the RF is highly irregular as demonstrated in Figure 1 where RGCs from macaque monkeys are shown to closely interlock and span the entire area of the visual window. Derived models which accurately describe this relationship progress our functional understanding of the retina and inspire future image processing research [8]. In fact, biologically inspired models of the retina, between stimulus and response, have been shown to outperform various machine vision techniques in terms of speed, power and performance [9].

Modelling of these temporal neural recordings, however, is challenging due to insufficient knowledge about the internal structure and interconnections between cells. Linear-nonlinear (LN) cascades are a popular class of quantitative models used

to describe the stimulus–response relationship [10]. In particular, LN models have been used to describe the processing in the retina [11] though the main drawback is that they lack any relationship between the derived parameters and underlying biophysics of the system [10]. System identification tools are useful in this case as they are suited to dynamical systems and allow for a better insight into the underlying physics of the biological system. First used to understand the responses of auditory neurons [12], output responses were recorded using white noise stimuli and inferences were made on mapping the stimulus to the response. As is often the case, white noise stimulation is preferred for modelling biological vision systems [13] as it remains controlled and is easily analysed mathematically. However, there is evidence that the use of artificial stimuli produces models that do not adequately describe responses to natural visual scenes [14]. Therefore, models created under these conditions using artificial stimuli may only be considered a subset of the full biological model under certain conditions.

Artificial Neural Network (ANN) methodologies, by definition, are designed to mimic biological aspects of the human brain [15] and through extension; the vision system. Specifically, Nonlinear Autoregressive Network with Exogenous Inputs (NARX) and k-Nearest Neighbours (kNN) approaches have been applied to neural encoding models in human vision [16] whilst methods, such as TDNN (Time Delayed Neural Network), MLP (Multi-layer Perceptron) and other ANN implementations have been used to derive models of retinal ganglion cell visual processing [17]–[19]. The NARMAX (nonlinear auto-regressive moving average with exogenous inputs) model [20]; a parametric system identification technique, which is a natural extension to NARX, has also been used within vision studies to model adaptation of photoreceptors to light in flies [21]. The NARMAX technique lends itself to a broad range of applications in several areas which include modelling robot behaviour [22], time series analysis [23], iceberg calving and detecting and tracking time-varying causality for EEG data [24]. In previous work, [18], [25] the NARMAX methodology has been utilised to help formulate a retina modelling development process and in particular, to express the biological input-output relationship using polynomial models.

In this work we expand on [25] by introducing, in addition to the NARMAX model, the self-organising fuzzy neural network (SOFNN) and NARX methodologies. The predictive performance of the investigated methodologies to adequately model a retinal ganglion cell’s output is evaluated. Performance is compared amongst these popular approaches, outlined in Section 2, with specific reference to the standard LN cascade technique. Section 3 provides details on the physiological experiments used for data collection and the methods utilised to pre-process the data to form an input-output time series configuration suitable for modelling. The results are presented in Section 4 where models have been derived based on two types of artificial stimuli. These models are then analysed further to determine any underlying system

dynamics. Finally, a concluding discussion based on the findings of this work is presented in Section 5 along with future directions for investigation.

II. METHODS

Deriving a quantitative relationship between stimulus and response of a RGC is challenging if we consider the internal cell structure that precedes them or the numerous interactions over the many interconnections between cells. To simplify this, we consider the problem with a black-box approach which aims to estimate a mathematical model for a regression dataset and apply a number of different methods to form this model. In keeping with traditional approaches, the Linear-Nonlinear (LN) cascade approach is also utilised as a comparison to the investigated approaches.

A. Linear-Nonlinear (LN)

The Linear-Nonlinear (LN) cascaded approach is a popular method of estimating the output firing rate of a neuron by applying the input to a linear temporal filter followed by a static non-linear transformation [10] and can be described by Eq. 1;

$$r(t) = F(a * S_t) \quad (1)$$

where a is the temporal linear filter, F is a static non-linearity and $a * S_t$ is the convolution of the temporal linear filter and stimulus S_t . The first step in estimating the response of the retina to visual stimuli is to compute the linear filter. This is typically accomplished by computing the spike triggered average (STA), which is simply the average stimulus preceding each spike (See Figure 2 for example).

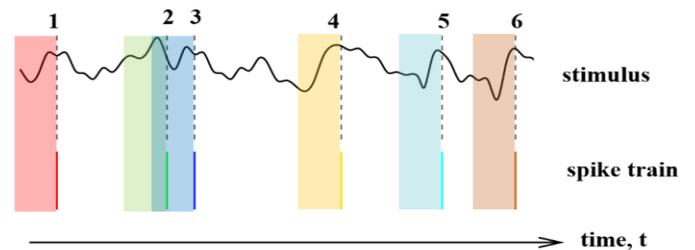


Figure 2: Calculating spike-triggered average

In [13], this is defined by Eq. 2 where T is the duration of the stimulus recording, $S_t f_t$ is the stimulus preceding a spike, f_t . Thus, the STA is the sum of all stimuli preceding a spike divided by the total number of spikes within the recording.

$$a = \frac{\sum_{t=1}^T S_t f_t}{\sum_{t=1}^T f_t} \quad (2)$$

The size of the temporal window is determined by examining the duration of the average response and ascertaining the point at which it converges to zero [13].

Determining the latter element of the LN cascade entails the convolution of the stimulus with the computed STA ($a * S_t$, Eq.1) and computing the static non-linearity (F). This is achieved by plotting the spike count as a function of the convolved stimulus and fitting a curve.

B. Nonlinear autoregressive exogenous model (NARX)

NARX (Nonlinear Autoregressive eXogenous) is part of the ANN family and is a model of a nonlinear neural network which accommodates dynamic inputs from a time series type dataset. It can learn to predict future values of the time series based on past information from the same time series, feedback input, and an additional time series referred to as the exogenous time series. Based on the same architecture as conventional recurrent neural networks, NARX provide a powerful solution to time series prediction that offers more effective learning and faster convergence over other ANNs [26]. A further advantage in principle is that one can use NARX networks, rather than conventional recurrent networks with complex differentiable nonlinearities, without any computational loss [27].

The topology of the network incorporates input, hidden and output processing element (PE) layers with the input to the network being fed by a number of delay units. Feedback from the output is also fed back to the hidden layer via delay units [28] as shown in Figure 3.

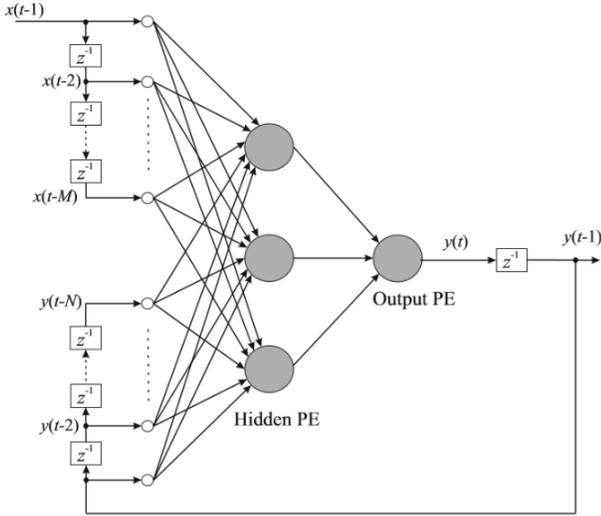


Figure 3: Architecture of a NARX network [28]

The description of the example model (Figure 3) can be denoted as:

$$y(t) = f\left(\sum_{i=1}^N a_i y(t-i) + \sum_{i=1}^M b_i x(t-i)\right), \quad (3)$$

where N, M, a_i and b_i are constants; $x(t)$ is the source input and $y(t)$ is the output of the network. Previously, NARX have been used to model various elements of the visual system including human tracking for robot vision applications [29] and the encoding of the natural visual system in humans through in vivo experimentation [16]. Although it has been proven that NARX is effective in its predictive performance of complex time series data [30], one of the disadvantages of models created via NARX is that they are not easily analysed due to their opaque nature in terms of the obtained mapping. This makes it very difficult to understand any underlying system dynamics that might otherwise be apparent in

alternative nonlinear system identification methods, for example NARMAX, which is discussed in the next section.

C. Nonlinear autoregressive moving average model with exogenous inputs (NARMAX)

A further improvement to the predictive powers of the NARX model can be achieved when the previous errors of the system are integrated as controlled variables [31]. A NARMAX model is formed as the result of this. The NARMAX approach is a popular system identification technique used when attempting to model the nonlinear relationship between the inputs and outputs (stimulus and response). It does this by representing the problem as a set of nonlinear difference equations. The NARMAX model, which is a natural extension of the ARMAX model [32], can be defined by:

$$y(t) = F[y(t-1), \dots, y(t-n_y), u(t-d), \dots, u(t-n_u), e(t-1), \dots, e(t-ne)] + e(t) \quad (4)$$

which accounts for the combined effects of noise, modelling errors and unmeasured disturbances concerning the inputs and outputs. Here, $u(t)$ and $y(t)$ are the input and output vectors respectively; $e(t)$ is system noise which is considered bounded and cannot be measured directly and n_y, n_u are the max output and input delays respectively. $F[\cdot]$, which is an unknown nonlinear function, is typically taken to be a polynomial expansion of the arguments.

To develop a NARMAX model, the structure of the nonlinear equation must first be identified along with the estimation of its parameters. The overall approach is made up of the following steps [24]:

- 1) Structure Detection: determine the terms within the model.
- 2) Parameter Estimation: tune the coefficients.
- 3) Model Validation: analyse model to avoid overfitting.
- 4) Prediction: output of the model at a future point in time.
- 5) Analysis: analyse model performance and determine the underlying dynamics of the system.

As the structure is typically unknown prior to the implementation, a range of possibilities exist to approximate the function including polynomial, rational and various ANN implementations [32], such as the NARX network. The polynomial models however offer the most attractive implementation with regards to visual modelling as they allow for the underlying dynamical properties of the system to be revealed and analysed. One solution to determine the important terms of the model can be achieved using an orthogonal least squares approach by computing the contribution that each potential model term makes to the system output. Building the system this way, term by term, exposes the significance of each new term added and allows for the avoidance of overfitting due to an excessive use of time lags or nonlinear function approximations [32] by ensuring that the model is as simple as possible and contains good generalisation properties. This approach simulates investigative modelling techniques where the important model

terms are introduced first and then the model is refined by adding in less significant terms. The only difference is that in the NARMAX method, the model terms can be identified directly from the data set. The unknown parameters and system noise can then be estimated and accommodated within the model. These procedures are now well established and have been used in many modelling domains [33].

D. Self-Organising Fuzzy Neural Network (SOFNN)

Another method which can be utilised to model and analyse time series type datasets is the Self-Organising Fuzzy Neural Network (SOFNN). A SOFNN is a hybrid network which has the capability to model and forecast a complex nonlinear system. It is capable of self-organising its architecture by adding and pruning neurons as required based on the complexity of the dataset. This alleviates the requirement of predetermining the model structure and estimation of the model parameters as the SOFNN can accomplish this without any in-depth knowledge of neural networks or fuzzy systems. The SOFNN approach has demonstrated good performance in applications of function approximation, complex system identification and time series prediction, further details of which can be found in [34]–[37].

The main architecture of the SOFNN is a five layer fuzzy neural network as depicted in Figure 4. These include an input layer, EBF (ellipsoidal basis function) layer, normalized layer, weighted layer and output layer. The SOFNN has the ability to reorganise the connections between these layers during the learning process. In the EBF layer, each neuron is a T -norm of Gaussian membership function (MF) attributed to the networks inputs (see Figure 5) where each neuron signifies the if-part of the fuzzy rule. The output from this layer is computed by products of the membership values of each input. The output of the EBF layer is normalised by the third layer, which contains an equal number of neurons, by dividing each output by the sum of all outputs.

The fourth network layer of the network is the weighted layer and signifies the consequent then-part of the fuzzy rules. Each neuron in this layer has two inputs, one of which is directly related to the output of the previous layer whilst the other is fed by a weighted bias. The product of these two inputs translate as the output to the final layer which contains a single neuron representing the summation of all incoming signals.

During the learning process of the SOFNN, its internal structure is dynamically modified through adding and pruning of neurons within the EBF layer to achieve an economical network size. Before adding a neuron to the network, existing membership functions are first examined to ascertain whether or not they can be modified to accommodate the new training sample while considering the generalisation performance of the overall network. This is determined using the following error criteria:

$$|\epsilon(t)| = |d_t - y_t|, \quad (5)$$

where d_t is the desired output of the system and y_t is the network output. If this error is greater than some user defined

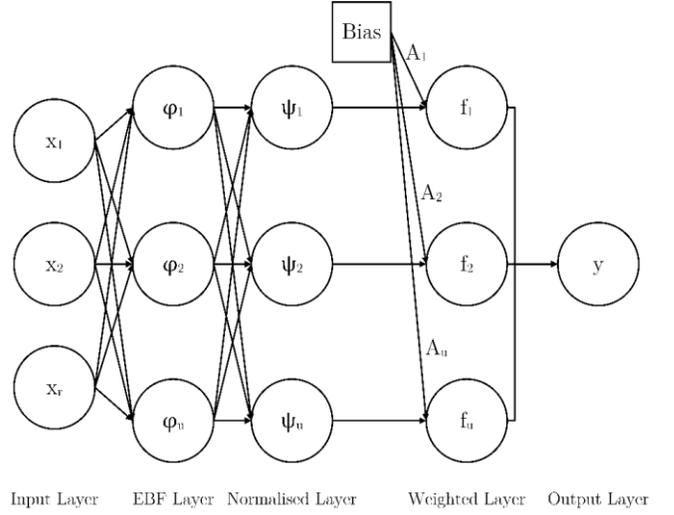


Figure 4: Layer Structure of SOFNN

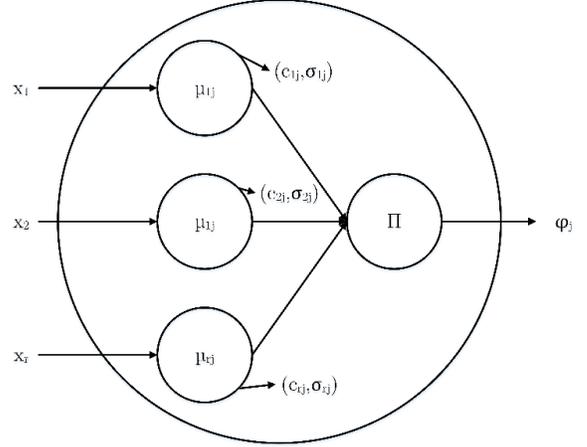


Figure 5: Internal Structure of EBF Neuron

threshold, δ , adding a new EBF neuron to the network will be considered, otherwise an existing membership function may be modified so that it appropriately clusters the new training sample.

Pruning of a neuron is governed by the importance of each neuron based on its contribution to the overall networks performance. The strategy is based on the optimal brain surgeon approach [38] which uses second derivative information to find the least important neuron and prune it from the network. If the subsequent performance of the SOFNN remains unchanged, the neuron is permanently deleted. Consequently, the neuron is restored should the performance be significantly degraded. An in depth explanation of the adding and pruning strategy is outlined in [36].

III. STIMULUS AND DATA PRE-PROCESSING

Neuronal data were recorded from retinas, which were isolated from dark adapted adult axolotl tiger salamanders, similar to the approach in [4], [39], where the retina is divided in half, with each half placed cell side down onto a multi-electrode array. Each image was projected onto the RGCs by a miniature organic light-emitting diode (OLED) display with

white light. A lens then de-magnifies the image and focuses it onto the photoreceptor layer of the isolated retina. The stimulus display ran at 60 Hz whilst the stimulus itself was updated at 30Hz, meaning a new stimulus presentation was made approximately every $33\frac{1}{3}$ ms. The neural responses (spikes) were recorded at 10kHz and binned at the stimulus update rate; meaning that all spikes that occur within the stimulus presentation timeframe are summed. Recorded spikes were sorted off-line by a cluster analysis of their shapes, and spike times were measured relative to the beginning of the stimulus presentation.

A. Stimulus

These recordings were performed while under stimulation using temporal and spatio-temporal Gaussian white noise sequences. Artificial white noise sequences are frequently utilised when determining various characteristics of RGCs, including the STA (Section 2.1), as this avoids cell adaptation to sustained stimuli, is relatively robust and spans a wide range of visual inputs [13]. An example of the stimulus is presented in Figure 6, where each image in the temporal sequence is presented sequentially. Figure 6(a) shows a set of images drawn from a randomly distributed Gaussian white noise sequence, used for full field illumination, where all pixels within each image are illuminated with the same light intensity, thus no spatial arrangement is observable. This is referred to as Full Field Flicker (FFF) and is the least complex form of artificial stimulus used within these experiments. Deriving models under these conditions, however, would only be relevant under a certain subset of conditions as the model would only consider the temporal component. Thus Figure 6(b) extends the stimulus input range to include a spatio-temporal input by introducing the binary checkerboard pattern. The Checker-Board Flicker (CBF), again drawn randomly from a Gaussian distribution, extends the complexity of the input due to the additional spatial component and is commonly used to determine characteristics of a cell's receptive field [40].

B. Data Pre-processing

The overall goal of the pre-processing stage is to manipulate the data so that they form a regression or classification dataset, i.e. input-output corresponding to the stimulus-response, which then can be used for developing the computational models. Recordings were supplied for a number of ganglion cells and organised within two datasets, both containing the visual stimuli and neural spike responses. The first dataset contained a large set of non-repeated stimuli (216000 samples for FFF and 258000 samples for CBF) that are suitable to ascertain characteristics such as the STA and to ensure that a sufficient number of varied stimuli are presented in order to evoke cell responses. The second dataset contained a much smaller set of stimuli (1200 samples) which were presented to the cells repeatedly.

Traditionally, only stimulus values within a cell's RF are considered for analysis as only values within this sensory space contribute to a cell's response. However, as FFF stimulus has uniform spatial intensity throughout, there is no need to extract the specific stimulus in the region of the RGCs

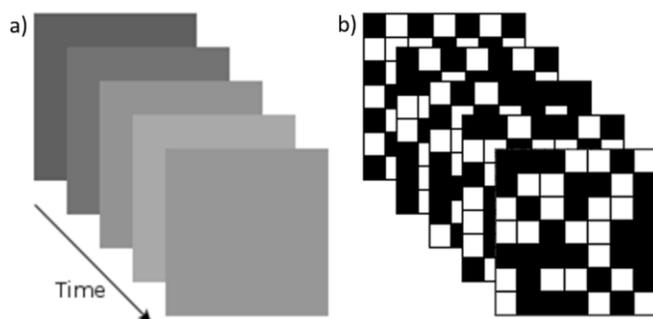


Figure 6: Pseudo-random sequence of a) Gaussian temporal sequence and b) spatio-temporal sequence.

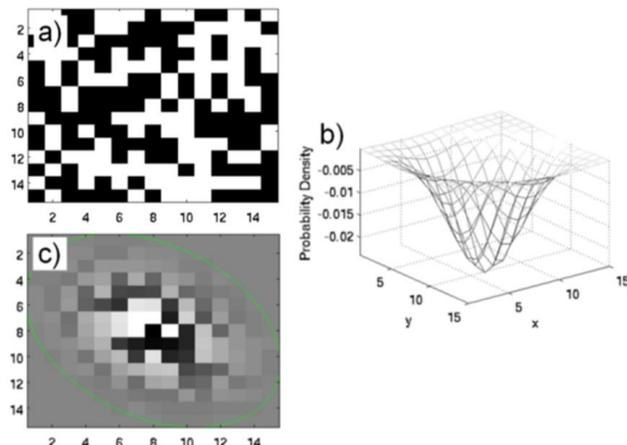


Figure 7: Pre-processing step which shows how the local stimulus pertaining to a cell's receptive field is weighted with a 2D Gaussian filter. (a) Local stimulus for a cell's receptive field. (b) 2D Gaussian used to weight the stimulus intensities. (c) Weighted image of the local stimulus intensities.

RF, thus the average intensity of each presented image is extracted instead. The neural response, originally recorded as a frequency of 10kHz, were binned at 30Hz to align well with the stimulus input forming a single input – output dataset.

As a pre-processing step for the CBF stimuli, the pertinent stimulus values must first be extracted from the checkerboard pattern (Figure 7(a)); here we extract only those checkerboard values located either inside, or on the border of the cell's RF. The RF is determined using a standard reverse-correlation method [4], [13] which is a technique for studying how sensory neurons summate signals from different times and locations to generate a response [41].

To emulate the processing that occurs between the photoreceptors and RGCs, the local stimulus within the RF is weighted using a 2D Gaussian filter (with a support of 3σ) [8], which is illustrated in Figure 7(b). From the resulting weighted stimulus, shown in Figure 7(c), the pixels within the region of the RF (green ellipse) are extracted and summed to form an input for the derived models. This results in a single value representing the CBF pattern for each time step rather than the individual pixel values (This is the standard approach [42] but, as it will be discussed later, the authors believe it merits further investigation). The binned neural response is again used as the output which is binned according to the stimulus update rate. In the case of the second dataset, i.e. the repeated trials, the mean of the binned spike rate is computed using the 43 trials and used as the model output.

IV. RESULTS

Recordings of the ganglion cell neural responses (spikes) to the FFF and CBF stimulation were provided for a number of different ganglion cells. Here we demonstrate analysis of two selected ganglion cells for each stimulus set, one *ON-cell* and one *OFF-cell*. The cell type is traditionally characterised by the shape of its temporal profile (STA) [13], [43], [44], whilst the length of the temporal window can be assessed by examining the duration of the average response and ascertaining the point at which it converges to zero [13]. Figure 8 illustrates the calculated profiles for both cells, where it was determined that 21 lagged values (700ms) of the time series were sufficient to capture the required behaviour. Figure 8(a) shows a temporal profile akin to what is described as a biphasic OFF type cell in [43]; we refer to this cell simply as an *OFF-cell* in this work. Figure 8(b) shows a profile that is typical of an *ON-cell*.

A. Temporal Artificial Stimuli

Determining the STA profile of the cell is also useful for indicating the number of lagged values to use for training the models. For instance, for results presented in this section, it was estimated that 21 lagged values would be sufficient to train each algorithm. Upon further investigation it was established that the use of 10 lagged values, essentially capturing the main STA characteristics (Figure 8(a)-(b)), was sufficient to train both the NARX and NARMAX methods and provided a marginal improvement in the estimated response. For the SOFNN method however, the full range of STA values worked best.

Results of the derived models for the FFF stimuli are presented in TABLE I and TABLE II respectively. For each of the different approaches model accuracy is measured using the root mean square error (RMSE) between the predicted and actual spike rate. From the results shown, it can be observed that the NARX method performs significantly better than the other investigated methods for both cells. Specifically, the performance increase of the NARX method over the LN method is quite substantial, with respect to the *OFF-cell* for both training and testing datasets. Surprisingly, integrating the previous errors of the system as controlled variables did not improve the predictive qualities of the model (Section II.C), as is evident by the NARMAX output.

Although the NARMAX approach achieved good results for the *OFF-cell* surpassing both the LN and SOFNN methods in performance, it did not improve upon the NARX model output which is less complex and requires considerably less computational power. Additionally, the SOFNN method, which has shown good performance in modelling output responses of isolated mice retinas [17], fails to provide an improved performance over the LN model for the salamander data.

To demonstrate the visible difference in performance of the NARX vs. LN method, the training and testing outputs are plotted in Figure 9 for 200 samples. The results presented here are for the *ON-cell*, which shows the performance of the NARX method to be improved in terms of the magnitude even though performance in terms of RMSE is not as significant in

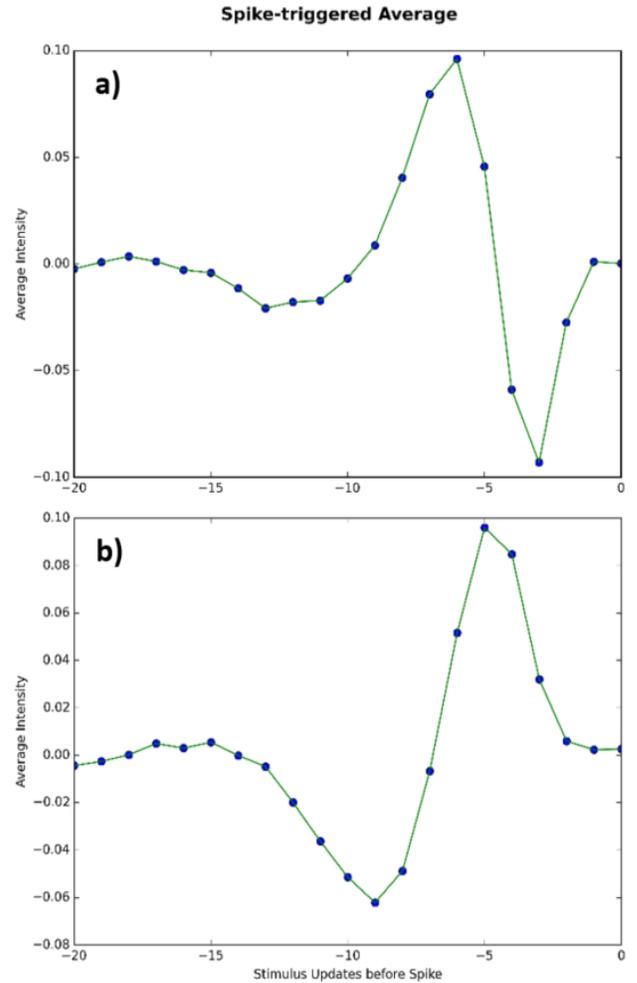


Figure 8: STA profile of (a) *OFF-cell* and (b) *ON-cell* using FFF stimuli

comparison to the *OFF-cell*. This is evident when comparing Figure 9(b) and Figure 9(d), which relates to the testing output. It can be observed that both methods perform well in terms of

TABLE I
RMSE VALUES FOR OFF-CELL USING FFF STIMULI

Model	Training RMSE	Testing RMSE
LN	0.71	0.63
NARMAX	0.67	0.61
NARX	0.52	0.49
SOFNN	0.77	0.68

TABLE II
RMSE VALUES FOR ON-CELL USING FFF STIMULI

Model	Training RMSE	Testing RMSE
LN	0.37	0.35
NARMAX	0.38	0.37
NARX	0.34	0.33
SOFNN	0.42	0.41

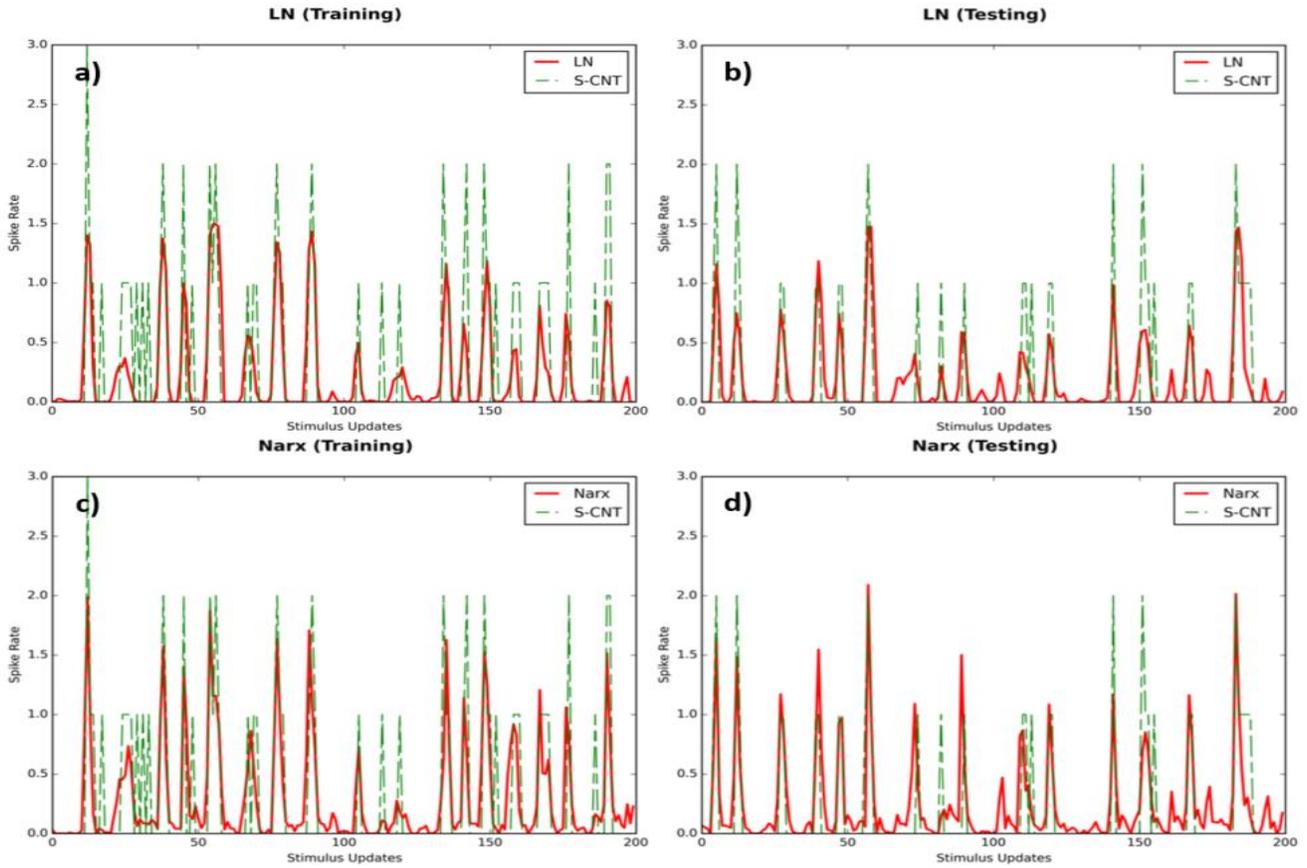


Figure 9: Prediction results of the ON-cell model using the LN and NARX models for FFF stimuli for: (a) the training samples of the LN method, (b) the testing samples of the LN method, (c) the training samples of the NARX method and (d) the testing samples of the NARX method.

predicting the timing of the spike rate though the NARX method additionally improves the magnitude of the predictions and in the majority of cases, reaches the target spike rate.

B. Spatio-Temporal Artificial Stimuli

To increase the complexity of the derived models, such that they can generalise over a more complex stimulus set, the CBF dataset is utilised. The results of the experiments for both datasets for the *OFF-cell* and *ON-cell* are outlined in TABLE III and TABLE IV respectively where model accuracy is measured in terms of the RMSE between the predicted and actual spike rate. Although the same cells are in use for these experiments, the stimulus sets differ and thus the results between datasets are not directly comparable. One observation, immediately noted, is that there is no clear separation between the performances of the LN approach vs. the other investigated methods. This is discussed later.

In terms of the RMSE values, the NARX method outperforms the other models for both cells during the training phase. Within the testing datasets however, the LN method performs on par with the NARX method for both testing sets with respect to the *ON-cell* and for the second dataset with respect to the *OFF-cell*. Among the remaining system identification models, the NARMAX model outperforms the SOFNN model for both the *OFF-cell* and *ON-cell*, with the exception of the SOFNN model achieving an equivalent performance on ‘Dataset 1’ for the *ON-cell*. Emphasis is

drawn to the fact that amongst the number of methods investigated, there is no significant improvement over the

TABLE III
RMSE VALUES FOR MODELS OF OFF-CELL USING CBF STIMULI

Model	Training RMSE (Dataset 1)	Testing RMSE (Dataset 1)	Testing RMSE (Dataset 2)
LN	0.35	0.35	0.27
NARMAX	0.35	0.36	0.28
NARX	0.34	0.35	0.27
SOFNN	0.36	0.37	0.30

TABLE IV
RMSE VALUES FOR MODELS OF ON-CELL USING CBF STIMULI

Model	Training RMSE (Dataset 1)	Testing RMSE (Dataset 1)	Testing RMSE (Dataset 2)
LN	0.38	0.38	0.24
NARMAX	0.39	0.38	0.25
NARX	0.37	0.37	0.24
SOFNN	0.39	0.38	0.27

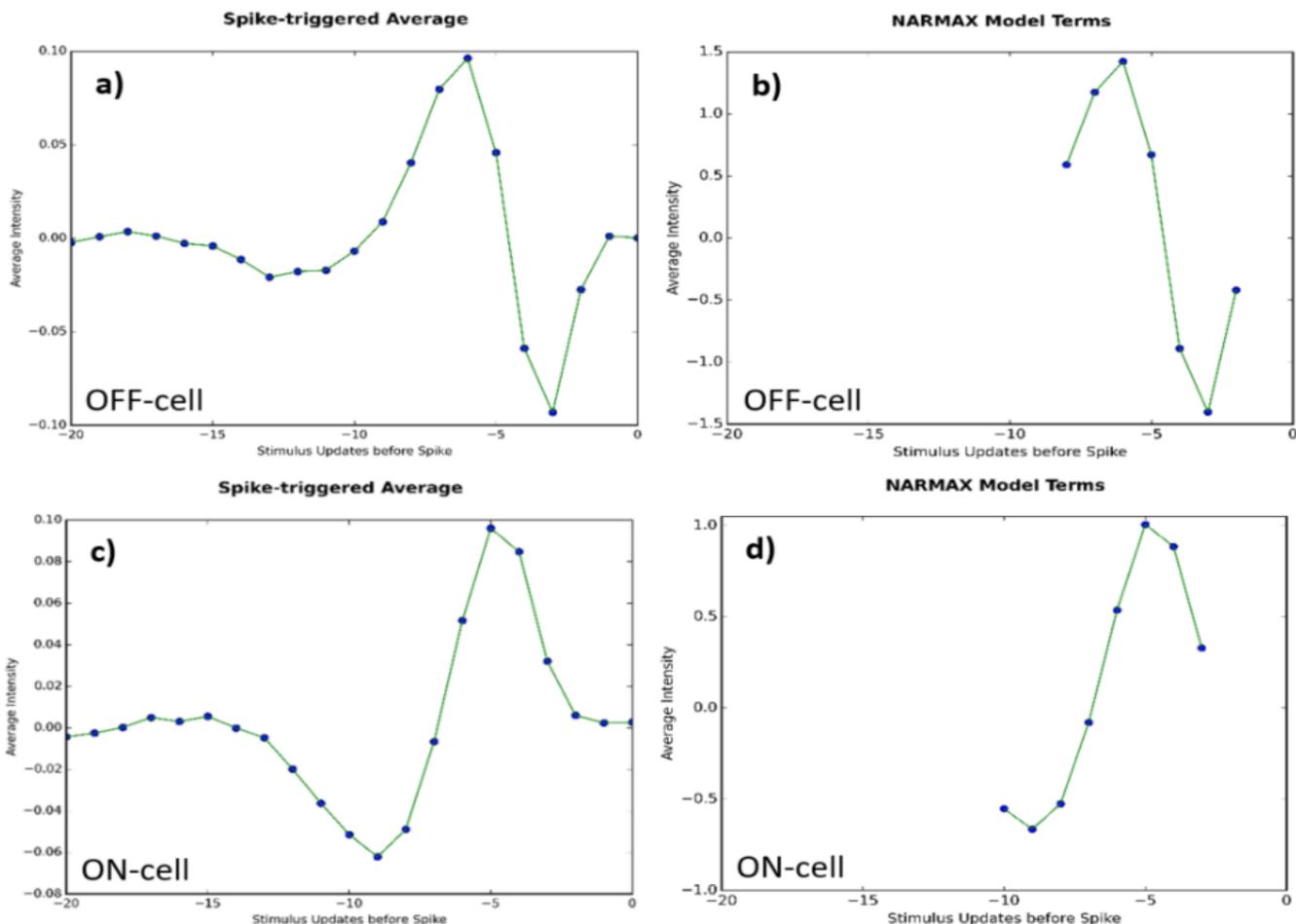


Figure 10: Plot illustrating STA and linear terms from cubic NARMAX model showing the (a) STA of the *OFF-cell*, (b) NARXMAX terms for the *OFF-cell* model, (c) STA of the *ON-cell* and (d) NARXMAX terms for the *ON-cell* model.

standard LN approach. We believe that, due to the increased spatial complexity of the stimulus, important information is being lost through the interpretation of the receptive field. This is currently achieved by extracting pertinent values inside the RF and simply summing or averaging to a single representative value. Interpreting the RF in this way disregards any spatial characteristics that may have proven to be important to the cells behaviour. This concept is explored further in Section V.

C. Model Analysis

The various models derived were analysed further to ascertain any underlying system dynamics that may be of interest to provide areas for further investigation. Models developed for both the FFF and CBF stimulus showed similar characteristics when under review thus here we report only on the analysis for the FFF stimulus set.

Analysis of the NARMAX model reveals some interesting observations within the model terms. To discuss further, we first compute the spike triggered average (STA) using the standard approach reported in [13]. The terms for each derived NARMAX model are then plotted and compared to the STA. Figure 10 illustrates the calculated STA and plotted NARMAX terms for both cells where the similarities between them are clearly observable. It is important to note that the NARMAX

terms are based on what the model deems as important when it is being derived. Therefore, for the *OFF-cell*, Figure 10(b) shows the terms considered most important when training the model which suggest that dramatic changes in the stimulus contrast levels are important. This is also the case when comparing the STA and terms for the *ON-cell* which are displayed in Figure 10(c) and Figure 10(d).

Due to the opaque nature of the NARX approach it is difficult to gain insight into any RGC models created using them. However, when analysing the weights of the input and hidden layers, a similarity can also be drawn with the calculated STA of the cell. Figure 11(a)-(b) shows this strong similarity when considering the most prominent neuron for the *OFF-cell*. Again, this is evident for the *ON-cell* illustrated in Figure 11(c)-(d).

Finally, the SOFNN technique allows us to gain some insight into the underlying dynamics of the data by analysing the fuzzy rules generated. For example, a rule generated by the SOFNN for the *OFF-cell* under FFF stimulation is as follows:

Rule 1: If Input 1 is A(-0.19113, 1.0375) AND Input 2 is A(0.15069, 1.0375) AND Input 3 is A(0.10993, 1.0375) AND Input 4 is A(-0.088315, 1.0375) AND Input 5 is A(0.30136, 1.0375) AND Input 6 is A(-0.085507, 1.0375) AND Input 7 is A(-0.064816, 1.0375) AND Input 8 is A(-0.08625, 1.0375)

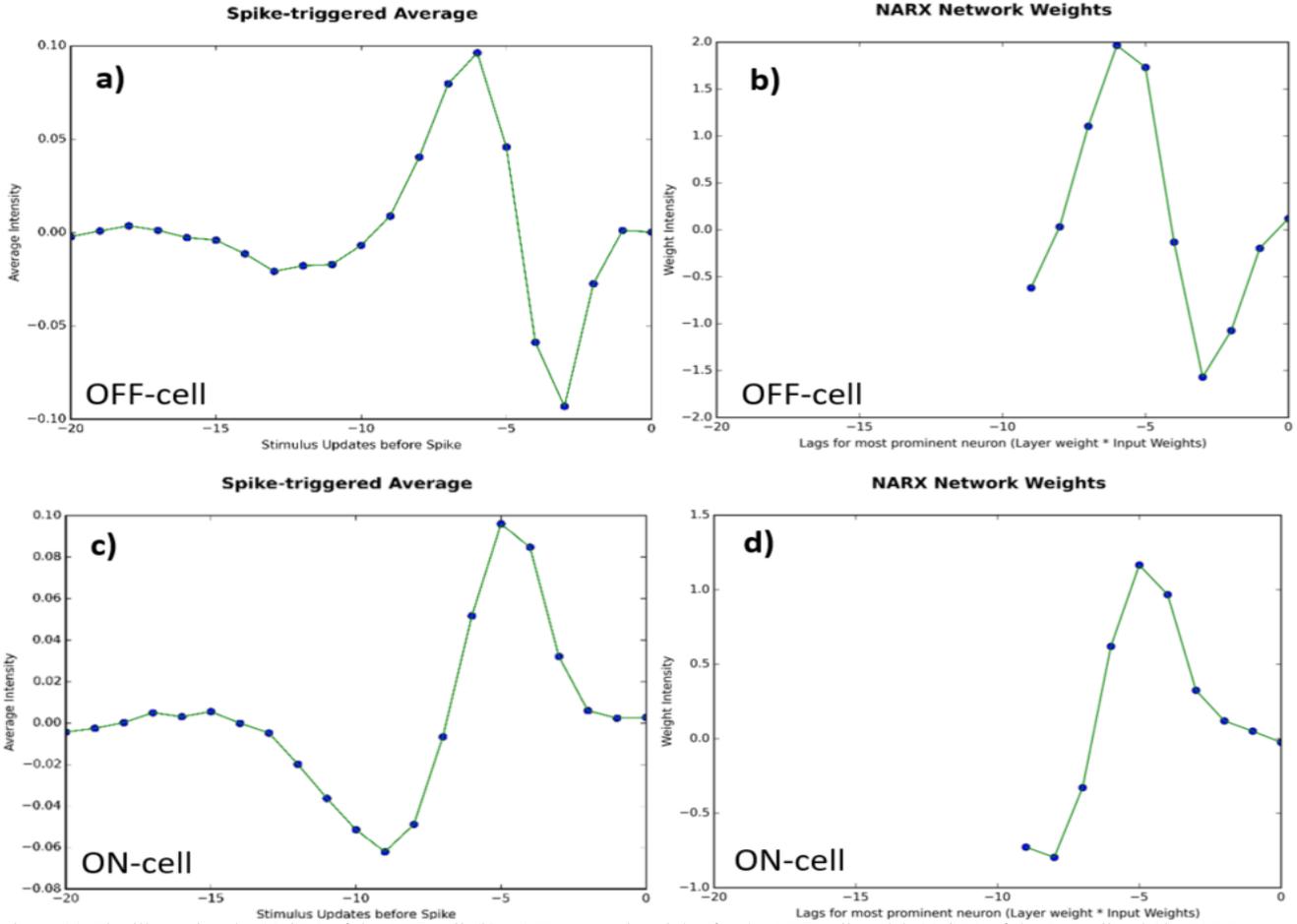


Figure 11: Plot illustrating the (a) STA of the *OFF-cell*, (b) NARX network weights for the *OFF-cell* model, (c) STA of the *ON-cell* and (d) NARX network weights for the *ON-cell* model.

AND Input 9 is A(0.23699, 1.0375) AND Input 10 is A(-0.12157, 1.0375) AND Input 11 is A(-0.71477, 0.96985) AND Input 12 is A(0.15307, 1.0375) AND Input 13 is A(0.15517, 1.0375) AND Input 14 is A(0.16102, 1.0375) AND Input 15 is A(0.16543, 1.0375) AND Input 16 is A(-0.010237, 1.0375) AND Input 17 is A(0.074938, 1.0375) AND Input 18 is A(0.1176, 1.0375) AND Input 19 is A(0.00056977, 1.0375) AND Input 20 is A(-0.20507, 1.0375) THEN

Output is $0.33452 + -0.0070681 * \text{Input } 1 + 0.0029931 * \text{Input } 2 + -0.42334 * \text{Input } 3 + -1.4056 * \text{Input } 4 + -0.89218 * \text{Input } 5 + 0.67233 * \text{Input } 6 + 1.428 * \text{Input } 7 + 1.1801 * \text{Input } 8 + 0.59511 * \text{Input } 9 + 0.12145 * \text{Input } 10 + -0.10781 * \text{Input } 11 + -0.26965 * \text{Input } 12 + -0.29999 * \text{Input } 13 + -0.34188 * \text{Input } 14 + -0.17775 * \text{Input } 15 + -0.053445 * \text{Input } 16 + -0.034553 * \text{Input } 17 + 0.018151 * \text{Input } 18 + 0.050873 * \text{Input } 19 + 0.01101 * \text{Input } 20.$

where A (centre, width) describes the membership function for each input. Similar to the machine learning models, the inputs to the SOFNN are lagged values of the stimulus sequence, so Input 1 corresponds to the current value of the series (i.e. zero milliseconds delay), and Input 20 corresponds to the value nineteen time steps in the past (i.e. 700ms delay). The result of

plotting the coefficients of the consequent part of the rule is shown in Figure 12(b). Remarkably, this approximately resembles the STA of the *OFF-cell*, which is shown in Figure 12(a) for comparison. Analysing the rules for the *ON-cell* yields a similar outcome (Figure 12(c) and Figure 12(d)). We hypothesise that the SOFNN, similar to the previous two methods, has identified the temporal characteristics of the STA as being the most important attribute of the stimulus that contribute to a cells response though further work in this area is ongoing. Both cells were adequately modelled by one neuron, corresponding to 1 fuzzy rule. Under CBF stimulation, the SOFNN model consisted of two neurons and consequently two fuzzy rules to model each cell. The increased network size, i.e. number of neurons, indicates that CBF stimuli are more complex to model than the FFF stimuli.

V. DISCUSSION AND FUTURE WORK

Modelling retinal ganglion cells within the retina is difficult due to insufficient knowledge about the internal components, their organisation and the complexity of the interactions within the system. Existing computational models are traditionally derived by quantitatively fitting particular sets of physiological data using an input-output analysis involving computational combinations of linear and nonlinear models

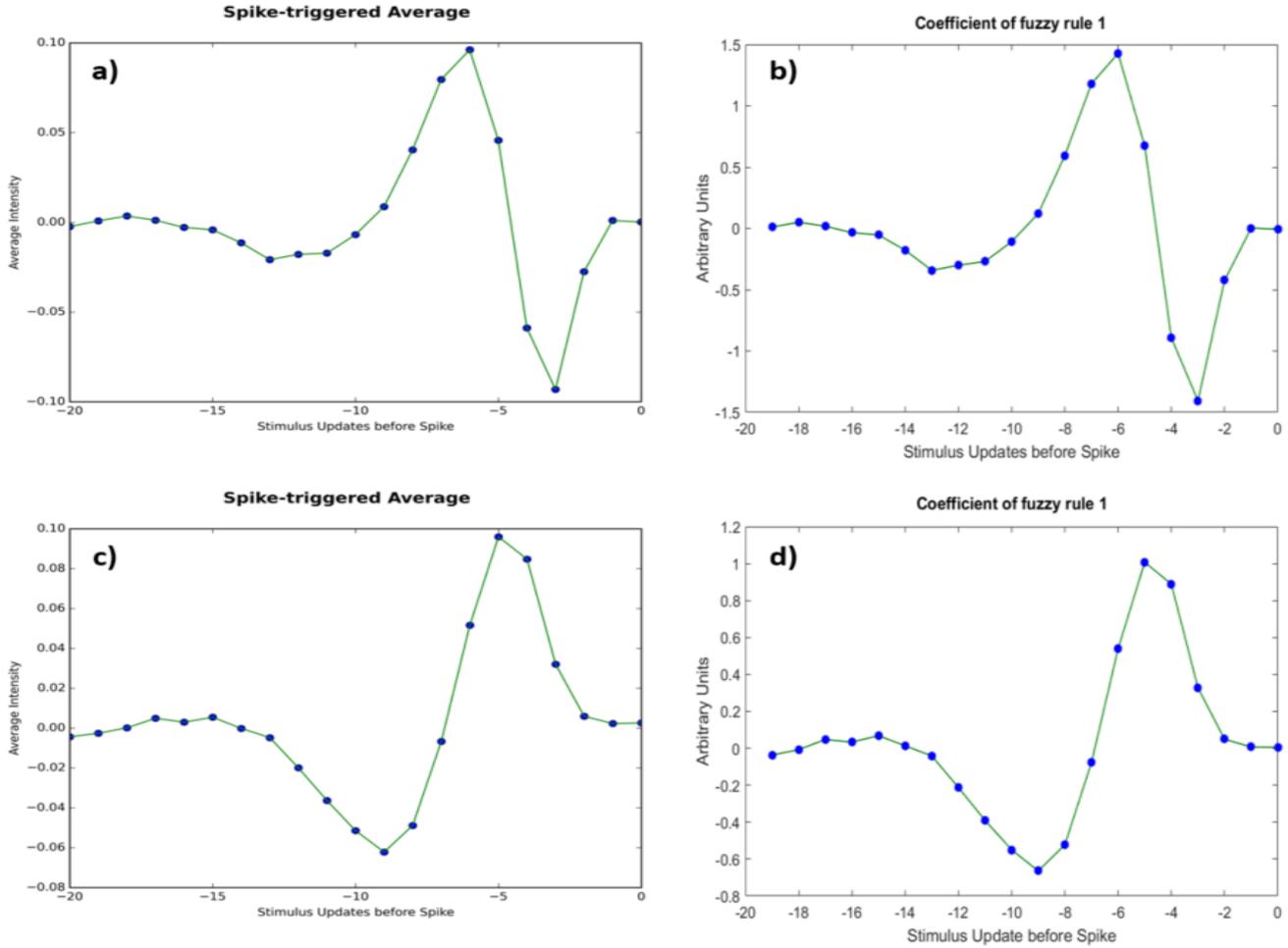


Figure 12: Plot illustrating the (a) STA of the *OFF-cell*, (b) Coefficients of fuzzy rule for *OFF-cell* model, (c) STA of the *ON-cell* and (d) Coefficients of fuzzy rule for *ON-cell* model.

that are generally complex and lack any relevance to the underlying biophysics. The work outlined in this paper explores the application and feasibility of modelling RGC's with system identification techniques as an alternative to the traditional linear-nonlinear approach. We present results based on the application of a selection of system identification techniques, namely NARX, NARMAX and SOFNN, to both temporal and spatio-temporal data revealing any underlying system dynamics that are observed after the modelling process.

The full-field temporal stimulus presented in Section 3.1 was the least complex stimulus considered for the work and consequently the explored models showed good performance in predicting the relationship between stimulus and response. In particular, the NARX method outperformed the other techniques in modelling both the *OFF-cell* and *ON-cell*. This performance increase is also clearly observable when reviewing the model for the *ON-cell* which shows the least difference in terms of the RMSE when compared to the LN model. Surprisingly, the NARMAX method did not offer a better performance with its increased complexity through integration of previous errors as controlled variables. Although the results show that the NARMAX outperformed

both the SOFNN and LN techniques, it was not able to outperform NARX which is significantly less complicated. The SOFNN did not perform as favourably in modelling the salamander RGCs but has shown good performance in past applications with mouse RGCs [17]. Its ability to capture characteristics that align well with the RGCs STA surpasses the representation observed through the NARMAX polynomial terms or the NARX internal layer weights though for this application it seems more appropriate to choose the simpler models that can generalise well over the input data, such as the NARX model.

With these interesting results for the temporal stimulus, we extended these modelling approaches to a more complex spatio-temporal stimulus (outlined in Section 3.1). The spatio-temporal artificial stimuli increased the complexity of the stimulus as it introduced the need to process the receptive field information pertaining to each cell by extracting pixels within the region of interest, weighting with a Gaussian filter and summing the result. Of the methods investigated to model the relationship of the increased complexity between the input and output, the NARX method again performed favourably in comparison to the other methods investigated. However, where the NARX clearly performed better for the FFF dataset,

the difference observed between the NARX and LN within the CBF dataset was diminished in terms of the RMSE. Here, the NARX had an improved performance with respect to the training dataset for both cells but performed on par with the LN method for the testing sets. The NARMAX method provided a slightly improved performance over the SOFNN method for the *OFF-cell* but an equal performance for the *ON-cell*. Similarly with the temporal data, characteristics akin to the STA of each cell were observable in all the system identification methods presented for the CBF data. The readability of such characteristics offers an advantage over more opaque approaches like LN that may provide a more in depth understanding of the underlying dynamics of the system, however further investigation into the relationship between these characteristics and the STA is ongoing.

Although the models presented adequately fit the real neural response, specifically the NARX method, there is quite a significant difference between the results for the temporal and spatio-temporal datasets. While not directly comparable, it can be observed that within the temporal data results, there is a clear separation between the performances of the standard LN approach vs the more bio-inspired techniques. Analysis of the spatio-temporal results revealed results which were not as clearly discriminable.

A. Future Work

Stemming from the comparative analysis between the temporal and spatio-temporal modelling approaches, we further investigated various aspects of the spatio-temporal modelling process. In particular, we questioned the transformation process of the receptive field and queried whether summing all of the spatial information to a singular representative value is sufficient enough to model an RGC efficiently. Given the increased complexity of the spatio-temporal to that of the temporal dataset it was hypothesised that summing the Gaussian weighted data within each receptive field was resulting in the significant loss of spatial information which could account for the increased complexity of fitting a model to the neural response. In [45] a method for retaining the spatial information is presented which calculates the STA spatially as well as temporally, filtering the stimulus with spatial information to create the input stimulus to a model. In this method, it is the LN method which benefits from this approach thus our initial investigation lead to constructing the linear filter within the LN approach from a spatial STA analysis.

The results obtained from this approach were marginally better in terms of RMSE and the magnitude of the nonlinear estimate; however the associated computational cost would be extremely large for the NARX and other bio-inspired methods without reducing the input space. This warrants further investigation into how the data within the receptive field can be compressed so that its influence is calculated correctly and with efficient computational complexity.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ICT-2011.9.11) under grant number [600954]

(“VISUALISE”). The experimental data contributing to this study have been supplied by the “Sensory Processing in the Retina” research group at the Department of Ophthalmology, University of Göttingen as part of the VISUALISE project.

References

- [1] C. I. Baker, “Handbook of Psychology, Behavioral Neuroscience, Second Edition,” vol. 3, Weiner, I.B., Nelson, R.J. and Mizumori, S., Ed. Wiley, 2012.
- [2] D. H. Hubel, *Eye, brain, and vision*, vol. 22. Scientific American Library New York, 1988.
- [3] T. Gollisch and M. Meister, “Eye smarter than scientists believed: neural computations in circuits of the retina.,” *Neuron*, vol. 65, no. 2, pp. 150–64, 2010.
- [4] T. Gollisch and M. Meister, “Rapid neural coding in the retina with relative spike latencies,” *Science (80-.)*, vol. 319, no. 5866, pp. 1108–1111, 2008.
- [5] W. F. Heine and C. L. Passaglia, “Spatial receptive field properties of rat retinal ganglion cells,” *Visual neuroscience*, vol. 28, no. 05, pp. 403–417, 2011.
- [6] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, “Spatio-temporal correlations and visual signalling in a complete neuronal population.,” *Nature*, vol. 454, no. 7207, pp. 995–9, 2008.
- [7] B. P. Olveczky, S. A. Baccus, and M. Meister, “Segregation of object and background motion in the retina.,” *Nature*, vol. 423, no. 6938, pp. 401–8, 2003.
- [8] D. Kerr, T. McGinnity, S. Coleman, and M. Clogenson, “A biologically inspired spiking model of visual processing for image feature detection,” *Neurocomputing*, vol. 158, pp. 268–280, 2015.
- [9] S. Shah and M. D. Levine, “Visual information processing in primate cone pathways. I. A model,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 26, no. 2, pp. 259–274, 1996.
- [10] S. Ostojic and N. Brunel, “From spiking neuron models to linear-nonlinear models,” *PLoS Comput. Biol.*, vol. 7, no. 1, p. e1001056, 2011.
- [11] J. W. Pillow, L. Paninski, V. J. Uzzell, E. P. Simoncelli, and E. Chichilnisky, “Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model,” *The Journal of Neuroscience*, vol. 25, no. 47, pp. 11003–11013, 2005.
- [12] E. De Boer and P. Kuyper, “Triggered correlation,” *Biomedical Engineering, IEEE Transactions on*, no. 3, pp. 169–179, 1968.
- [13] E. J. Chichilnisky, “A simple white noise analysis of neuronal light responses.,” *Network*, vol. 12, no. 2, pp. 199–213, 2001.
- [14] C. Kayser, K. P. Körding, and P. König, “Processing of complex stimuli and natural scenes in the visual cortex,” *Current opinion in neurobiology*, vol. 14, no. 4, pp. 468–473, 2004.
- [15] D. J. Livingstone, *Artificial Neural Networks: Methods and Applications (Methods in Molecular Biology)*. Humana Press, 2008.

- [16] N. P. P. A. Sousa, "Neural Encoding Models in Natural Vision," 2013.
- [17] S. McDonald, D. Kerr, S. Coleman, P. Vance, and T. McGinnity, "Modelling retinal ganglion cells using self-organising fuzzy neural networks," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, 2015, pp. 1–8.
- [18] D. Kerr, S. Coleman, and T. McGinnity, "Modelling and Analysis of Retinal Ganglion Cells with Neural Networks," *Irish Machine Vision and Image Processing*, 2014.
- [19] G. P. Das, P. Vance, D. Kerr, S. A. Coleman, and T. M. McGinnity, "Modelling Retinal Ganglion Cells Stimulated with Static Natural Images," 2016, p. (in press).
- [20] S. Billings and W. Voon, "Least squares parameter estimation algorithms for non-linear systems," 1984.
- [21] U. Friederich, D. Coca, S. Billings, and M. Juusola, "Data modelling for analysis of adaptive changes in fly photoreceptors," in *Neural Information Processing*, 2009, pp. 34–48.
- [22] D. Kerr, U. Nehmzow, and S. Billings, "Towards automated code generation for autonomous mobile robots," 2010.
- [23] Y. Gao and M. J. Er, "NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches," *Fuzzy Sets and Systems*, vol. 150, no. 2, pp. 331–350, 2005.
- [24] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.
- [25] D. Kerr, M. McGinnity, and S. Coleman, "Modelling and Analysis of Retinal Ganglion Cells Through System Identification," 2014.
- [26] L. Iliadis, I. Maglogiannis, and H. Papadopoulos, *Artificial Intelligence Applications and Innovations: 10th IFIP WG 12.5 International Conference, AIAI 2014, Rhodes, Greece, September 19-21, 2014, Proceedings*, vol. 436. Springer, 2014.
- [27] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent NARX neural networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 27, no. 2, pp. 208–215, 1997.
- [28] M. Bianchini, M. Maggini, and L. C. Jain, *Handbook on Neural Information Processing*. Springer, 2013.
- [29] E. Petrovic, Z. Cojbasic, D. Ristic-Durrant, V. Nikolic, I. Ciric, and S. Matic, "Kalman Filter and NARX neural network for robot vision based human tracking," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 12, no. 1, pp. 43–51, 2013.
- [30] H. Xie, H. Tang, and Y.-H. Liao, "Time series prediction based on NARX neural networks: An advanced approach," in *Machine Learning and Cybernetics, 2009 International Conference on*, 2009, vol. 3, pp. 1275–1279.
- [31] A. Hornstein and U. Parlitz, "Bias reduction for time series models based on support vector regression," *International Journal of Bifurcation and Chaos*, vol. 14, no. 06, pp. 1947–1956, 2004.
- [32] S. A. Billings and D. Coca, "Identification of NARMAX and related models," *Research Report-University of Sheffield, Department of Automatic Control and Systems Engineering*, 2001.
- [33] S. Chen and S. Billings, "Representations of non-linear systems: the NARMAX model," *International Journal of Control*, vol. 49, no. 3, pp. 1013–1032, 1989.
- [34] G. Leng, G. Prasad, and T. McGinnity, "A new approach to generate a self-organizing fuzzy neural network model," in *Systems, Man and Cybernetics, 2002 IEEE International Conference on*, 2002, vol. 4, p. 6–pp.
- [35] G. Leng, T. M. McGinnity, and G. Prasad, "Design for self-organizing fuzzy neural networks based on genetic algorithms," *Fuzzy Systems, IEEE Transactions on*, vol. 14, no. 6, pp. 755–766, 2006.
- [36] G. Leng, T. M. McGinnity, and G. Prasad, "An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network," *Fuzzy sets and systems*, vol. 150, no. 2, pp. 211–243, 2005.
- [37] G. Leng, G. Prasad, and T. M. McGinnity, "An on-line algorithm for creating self-organizing fuzzy neural networks," *Neural Networks*, vol. 17, no. 10, pp. 1477–1493, 2004.
- [38] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 1, pp. 116–132, 1985.
- [39] J. K. Liu and T. Gollisch, "Spike-Triggered Covariance Analysis Reveals Phenomenological Diversity of Contrast Adaptation in the Retina," *PLoS Comput Biol*, vol. 11, no. 7, p. e1004425, 2015.
- [40] T. Gollisch and M. Meister, "Modeling convergent ON and OFF pathways in the early visual system.," *Biol Cybern*, vol. 99, no. 4–5, pp. 263–78, 2008.
- [41] D. Ringach and R. Shapley, "Reverse correlation in neurophysiology," *Cognitive Science*, vol. 28, no. 2, pp. 147–166, 2004.
- [42] G. W. Schwartz, H. Okawa, F. A. Dunn, J. L. Morgan, D. Kerschensteiner, R. O. Wong, and F. Rieke, "The spatial structure of a nonlinear receptive field.," *Nat. Neurosci.*, vol. 15, no. 11, pp. 1572–80, 2012.
- [43] R. Segev, J. Puchalla, and M. J. Berry, "Functional organization of ganglion cells in the salamander retina.," *J. Neurophysiol.*, vol. 95, no. 4, pp. 2277–92, 2006.
- [44] D. R. Cantrell, J. Cang, J. B. Troy, and X. Liu, "Non-centered spike-triggered covariance analysis reveals neurotrophin-3 as a developmental regulator of receptive field properties of ON-OFF retinal ganglion cells," *PLoS Comput Biol*, vol. 6, no. 10, pp. e1000967–e1000967, 2010.
- [45] O. Schwartz, J. W. Pillow, N. C. Rust, and E. P. Simoncelli, "Spike-triggered neural characterization," *Journal of Vision*, vol. 6, no. 4, p. 13, 2006.