A Simple Parametric Representation of the Hodgkin-Huxley Model

Alejandro Rodríguez-Collado^{1*} , Cristina Rueda¹

¹ Department of Statistics and Operations Research, Universidad de Valladolid, Valladolid, Spain.

These authors contributed equally to this work.

* alejandrorodriguezcollado@gmail.com

Abstract

The Hodgkin-Huxley model, decades after its first presentation, is still a reference model in neuroscience as it has successfully reproduced the electrophysiological activity of many organisms. The primary signal in the model represents the membrane potential of a neuron. A parametric and simple representation of this signal is presented in this paper.

The new proposal is an adapted Frequency Modulated Möbius multicomponent model defined as a flexible decomposition in waves that describe the signal morphology. A specific feature of the new model is that the parameters are subject to interpretable restrictions.

A broad simulation experiment is conducted to show the new model accurately represents the simulated Hodgkin-Huxley signal. Moreover, the model potential to predict the neuron's relevant characteristics, described with parameters of the Hodgkin Huxley model, is shown using different Machine Learning methods. The proposed model is also validated with real data from Squid Giant Axons. The comparison of the parameter configuration between the simulated and real data demonstrated the flexibility of the model as well as interesting differences.

Author summary

Alejandro Rodríguez-Collado. I received the double degree in Statistics and Computer Engineering and the Master's degree in Business Intelligence and Big Data from the Universidad de Valladolid in 2019 and 2020, respectively. I work as researcher and Professor for the Department of Statistics and Operational Research at the Universidad de Valladolid. My main research interests include oscillatory signal processing, neuroscience, multivariate data analysis and supervised learning.

Cristina Rueda. I received the BS degree in mathematics from the Universidad de Valladolid in 1987 and the PhD degree in statistical science from the Universidad de Valladolid in 1989. I am currently Professor in the Department of Statistics and Operational Research at the Universidad de Valladolid. My main research interests include statistical inference methods under restrictions, circular data, computational biology, and statistical methods for signal analysis.

Introduction

Neuroscience is an interdisciplinary science that studies the cellular, functional, behavioral, evolutionary, computational, molecular, and medical aspects of the nervous system. Many specialists from different areas of knowledge, such as physicists, chemists, mathematicians, computer engineers, and psychologists, have contributed to the field research. The mathematical approach is one of the most preferred ones, particularly in studying the electrophysiological activity between neurons. The signal that received most of the attention is the neuron membrane potential, which is the difference in electric charge between the cell's interior and exterior. This signal is composed of various Action Potential Curves (APs). A single AP lasts a few milliseconds and consists of 3 stages: depolarization, repolarization, and hyperpolarization. For researchers, APs are of special importance: they are the informational unit between neurons, and their number and shape determine the morphological, functional, and genetic profile of the cell. For more detail see, [1], [2], [3], [4] or [5].

According to their dynamic behavior, neurons can be classified into excitable (generate an individual AP) or oscillatory (generate repetitive APs). Neurons that do not generate APs are called non-excitable. Neuronal cells of a similar type habitually exhibit identical behaviors. For instance, cardiac myocytes are usually oscillatory, while cortical neurons are mostly excitable. Furthermore, oscillatory neurons have been sub-classified in this work by the observed number of APs. A signal where several APs are observed is known as a Spike Train.

The most broadly considered mathematical model for describing AP dynamics is the Hodgkin-Huxley (HH) model, presented in [6]. More than a half-century later, this model remains key in neuroscience due to its innovative concept of modeling neuronal dynamics as a system of Ordinary Differential Equations (ODE) and its accurate representation of the electrophysiological neuronal activity. Specifically, the neuron's membrane potential is stated to behave like an electrical circuit with three currents associated, each with three different types of ions. Sodium (Na^+), potassium (K^+) and another which is a non-specific leak current, mainly due to the influx of chlorine (Cl^-). However, the HH model lacks identifiability as various parameter configurations can lead to the same observed signal. In addition, the model is not robust as minor manipulations of the values of the parameters can change its output completely ([7], [8]).

Many models that have been developed afterward are either simplifications or extensions of the HH model. Some emulate the HH model as biophysically realistic, whereas others seek more simple models. Among the first group are the Hopfield model and the Van der Pol oscillator's extensions, such as the Fitz-Hugh Nagumo model. In the second group, some popular choices are the family of LIF models and the Izhikevich model. Some basic references about these models are [9], [10], [11], [12] and [13]. All the above are mechanistic models. The counterpart of the mechanistic models is the data-driven approach. Models based in data science, statistics and Machine Learning are in ever-rising popularity due to the increase in data availability and quality as noted in [14]. Our proposal can be framed within this class of models.

The Frequency Modulated Möbius (FMM) approach and others, such as the Fourier method, are encompassed in the AM-FM decompositions. A general overview of AM-FM decompositions and time-frequency signal analysis can be found in [15] and [16]. In particular, the FMM decomposition assumes a constant amplitude and a frequency that is modeled as a Möbius transformation. The monocomponent FMM model is presented in [17]. It shows how it accurately fits a wide variety of oscillatory patterns. The multicomponent FMM model is introduced in [18] and this concisely demonstrates its potential in neuroscience. Moreover, an exciting application for the automatic

analysis of electrocardiograms is presented in [19].

This paper's main goal is to show that the FMM model faithfully represents the AP signals derived from an HH model. To that end, a new FMM model is presented, denoted as FMM_{ST}, where ST stands for Spike Train. It is an FMM multicomponent model with restrictions on the parameters. Expressly, the model assumes that the Spike Train is the concatenation of a fixed number of successive spikes with the same shape, each one described with a bi-component FMM.

In order to validate the model, a simulated experiment has been designed. A total of 5000 HH signals, corresponding to a wide variety of parametric configurations, has been generated according to a factorial design for the most relevant HH parameters. It is shown that the estimated FMM_{ST} signals accurately predict the simulated signal across all the parameter configurations. Furthermore, using Machine Learning, the potential of the FMM parameters to predict relevant HH parameters is also shown.

On the other hand, the model's flexibility and good performance are also illustrated with the analysis of real data from the Squid Giant Axon Membrane Potential (SGAMP) database, originally from [20].

Methods

HH model

The presentation in this section follows those in [21] and [22]. The notation for HH variables and parameters is slightly changed from those used in these papers to avoid confusion with other terms introduced later in the paper.

The HH model is defined, see Definition 1 below, by a nonlinear ODE system for four state variables: the membrane potential (X) and the three opening probabilities of the ion gates (m, n, h). Furthermore, X depends on the input stimulus $I(t)$ generated by other neurons' post-synaptic currents. On their behalf, the variables m, n and h are referred to as voltage-gated channels as they depend on the membrane potential through the six ion flow rate functions denoted by a_j and b_j , $j \in \{m, n, h\}$, as explained in [4].

Definition 1. Hodgkin-Huxley (HH) model.

$$\begin{bmatrix} \frac{dX}{dt} \\ \frac{dm}{dt} \\ \frac{dn}{dt} \\ \frac{dh}{dt} \end{bmatrix} = \begin{bmatrix} \frac{1}{C}[-g_K n^4(X - V_K) - g_{Na} m^3 h(X - V_{Na}) - g_L(X - V_L) + I(t)] \\ (1 - m)a_m(X) - mb_m(X) \\ (1 - n)a_n(X) - nb_n(X) \\ (1 - h)a_h(X) - hb_h(X) \end{bmatrix}$$

The first equation in Definition 1 depends on the parameters of the cell capacitance (C), the maximum conductances (g_K, g_{Na}, g_L) and the equilibrium potentials (V_K, V_{Na}, V_L) of the ionic currents. The six ion flow rate functions increase the number of parameters to more than twenty parameters as explained in [22]. Hence, the parametric space of the model can be simplified by considering linear scalars of the ion flow rate functions ($\tilde{a}_m, \tilde{a}_n, \tilde{a}_h, \tilde{b}_m, \tilde{b}_n, \tilde{b}_h$), as done in [7] and [5].

An interesting simplification of the HH parametric space is the pair $(\mathcal{S}, \mathcal{K})$ that has been recently considered in [7] and is defined as follows:

$$\mathcal{S} = \frac{g_{Na}}{g_{Na} + g_K} \quad \mathcal{K} = \frac{\tilde{a}_n + \tilde{b}_m}{\tilde{a}_n + \tilde{b}_n + \tilde{a}_m + \tilde{b}_m}.$$

The authors of [7] explain the properties of the $(\mathcal{S}, \mathcal{K})$ pair. In particular, they claim that the neuron's excitability phenomenon is essentially bidimensional, being determined by the structure of the neuron as a cell and its ionic current kinetics. While $\mathcal{S}$ captures the neuron's structural information, $\mathcal{K}$ represents the kinetics of the ionic gates. Moreover, $(\mathcal{S}, \mathcal{K})$ are less sensitive to slight changes in the signal than the primary HH parameters. However, such a drastic reduction in dimensions does not give a complete representation of the model.

FMM approach

Let $t_i, i = 1, \dots, n$ denote the vector of observed time points and $X(t_i)$ the observed data, which in this paper is the potential difference in the neuron's membrane. It is assumed that the time points are in $[0, 2\pi)$. Otherwise, consider $t' \in [t_0, T + t_0]$ with t_0 as the initial time value and T as the period. Transform the time points by $t = \frac{(t' - t_0)2\pi}{T}$.

Let, $\mathbf{v} = (A, \alpha, \beta, \omega)'$ be the four-dimensional parameters describing a monocomponent FMM signal, defined as the following wave:
 $W(t, \mathbf{v}) = A \cos(\phi(t, \alpha, \beta, \omega))$, where A is amplitude and,

$$\phi(t, \alpha, \beta, \omega) = \beta + 2 \arctan(\omega \tan(\frac{t - \alpha}{2})) \quad (1)$$

is the phase. Where, α is a location parameter and the parameters β and ω determine the skewness and kurtosis. More details about the parameters can be found in [17].

The FMM approach relies upon a signal plus error model, as follows:

$$X(t_i) = \mu(t_i, \boldsymbol{\theta}) + e(t_i) = M + \sum_{j=1}^m W(t_i, \mathbf{v}_j) + e(t_i), \quad i = 1, \dots, n; \quad (2)$$

where, $\boldsymbol{\theta} = (M, \mathbf{v}_1, \dots, \mathbf{v}_m)$ and $(e(t_1), \dots, e(t_n))' \sim N_n(0, \sigma^2 \mathbf{I})$.

The papers [19] and [18] consider particular FMM models, show the broad type of signals that the model represents, provide properties, and interpret the parameters as well as detail the algorithm used to fit the models. In particular, in the second paper, its potential in neuroscience is concisely shown.

Depending on the application, the waves of the model represent different physiological processes. For instance, in the ECG case, the waves of the FMM model represent the five fundamental ECG upward and downward deflections, which are universally named P, QRS complex (a wave complex), and T.

The AP signals are modeled using two (or three in some cases) components, the first being much more relevant than the rest. This component, denominated Dominant Component, identifies when the neuron spikes, allows the AP's approximate reconstruction in the presence of noise, and the detection and identification of overlapping spikes. Moreover, theoretical properties are derived for the Dominant Component. All of these are shown in [18].

In Fig. 1, AP single signals and the FMM_{ST} predictions are shown. Fig. 1(A) illustrates the role of each parameter, while Fig. 1 (B) shows the wave composition for an excitatory and an oscillatory neuron.

The FMM_{ST} model is a particular multicomponent FMM model, where restrictions on the parameters are imposed and is defined as follows:

Fig 1. (A) Different FMM wave patterns generated by the parameters variation. **(B)** Components plots for the FMM₂ models for an AP from an excitable neuron (left) and from an oscillatory neuron (right) .

Definition 2. FMM_{ST} Model

For $i=1, \dots, n$:

$$X(t_i) = \mu(t_i, \theta) + e(t_i) = M + \sum_{S=1}^s W(t_i, \mathbf{v}_S^A) + W(t_i, \mathbf{v}_S^B) + e(t_i),$$

where,

- $\theta = (M, \mathbf{v}_1^A, \dots, \mathbf{v}_s^A, \mathbf{v}_1^B, \dots, \mathbf{v}_s^B)$
- $M \in \mathbb{R}; \mathbf{v}_S^A, \mathbf{v}_S^B \in \Theta = \mathbb{R}^+ \times [0, 2\pi] \times [0, 2\pi] \times [0, 1]; S = 1, \dots, s,$
- $A_S^A < A_S^B; S = 1, \dots, s$
- $\alpha_1^A \leq \alpha_1^B \leq \alpha_2^A \leq \alpha_2^B \dots \leq \alpha_s^A \leq \alpha_s^B \leq \alpha_1^A$
- $\beta_1^A = \beta_s^A$ and $\beta_1^B = \beta_s^B; S = 2, \dots, s.$
- $\omega_1^A = \omega_s^A$ and $\omega_1^B = \omega_s^B; S = 2, \dots, s.$
- $(e(t_1), \dots, e(t_n))' \sim N_n(0, \sigma^2 \mathbf{I}),$

where s is the number of spikes that is assumed known and can be easily estimated with a naive method, as is detailed in the estimation algorithm section.

The model also assumes that each spike is modeled with two waves with parameters $\mathbf{v}^A$ and $\mathbf{v}^B$, respectively. The restrictions on the A s and the α s guarantee identifiability, and the restrictions on the ω s and β s represent the assumption of spikes having the same shape.

The number of free parameters of the FMM_{ST} is $1 + 4s + 4$.

Additional restrictions may also be applied, depending on the application at hand. In particular, we will consider the following restrictions on the parameters A to analyse the SGAMP data:

$$A_1^A = A_S^A \text{ and } A_1^B = A_S^B; \quad S = 2, \dots, s \quad (3)$$

In this case, the number of free parameters is reduced to $1 + 2s + 6$.

Moreover, other parameters of interest in practice are the distance between the A and B component for a given AP and the distance between consecutive APs. These are defined as follows,

$$\begin{aligned} d_S^{AB} &= 1 - \cos(\alpha_S^A - \alpha_S^B) \quad S = 1, \dots, s \\ d_S^{AP} &= 1 - \cos(\alpha_S^A - \alpha_{S+1}^A) \quad S = 1, \dots, s - 1 \end{aligned}$$

In some applications, it would be fair to assume $d_1^{AB} = d_S^{AB}$; $S = 2, \dots, s$, which implies a reduction in the number of free parameters to $1 + s + 8$.

It would also be suitable to assume that the distances between consecutive spikes are constant, for instance in controlled experiments without stimulus changes. The restrictions that apply for that case are:

$$d_1^{AP} = d_S^{AP}; \quad S = 2, \dots, s - 1 \quad (4)$$

and the number of free parameters is then reduced to $1 + 1 + 8$.

Estimation algorithm

The implementation of FMM models in the programming language R, including applying the defined restrictions, is openly available in [23]. It is assumed that the segments to be analyzed represent complete spikes, in particular, $X(t_1) \simeq X(t_n)$. s is easily determined by a naive method based on a threshold as proposes [24]. This threshold is $k = 2.5\sigma_X$, σ_X being the sample standard deviation of the observed data.

Occasionally, two different parameter configurations represent a given signal equally well. However, one is physiologically more plausible. In that case, additional restrictions on the parameters are needed to guarantee that the solution is the one expected. In the case of SGAMP signals, it is assumed that a spike has a prominent Dominant Component and then that $A^A - A^B > C$, for a given threshold, C .

Validation measure

The goodness of fit of the model is measured with an R^2 statistic, which is the proportion of the variance explained by a model out of the total variance, as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (X(t_i) - \hat{\mu}(t_i, \boldsymbol{\theta}))^2}{\sum_{i=1}^n (X(t_i) - \bar{X})^2} \quad (5)$$

In this paper, $R_{\text{FMM}_{\text{ST}}}^2$, $R_{\text{FMM}_{\text{ST}^*}}^2$, $R_{\text{FMM}_{s^*}}^2$, $R_{\text{FMM}_{s^{**}}}^2$ refer to the R^2 value for the FMM_{ST} model, the FMM_{ST} model with restrictions on the A s, the FMM_s model with restrictions on β s and ω s, and the FMM_{s*} model with restrictions on the A s, respectively.

Machine Learning Supervised Methods

Several Machine Learning Supervised methods have been considered in the paper. At one end, the simple Linear Regression (LR) that serves as a benchmark approach. At the other extreme, the complex and “black box” Support Vector Machines of RBF Kernel (SVM) approach that has been proved to achieve excellent results in neuronal dynamics, as seen in [7] and [13], among others. Random Forest (RF) and Gradient Boosting Machines (GBM) are complex methods that provide interpretable results between the underfitting LR and the overfitting SVM, which have also been considered. [25] and [26] are essential references to learn about the procedures. The R packages [27], [28], [29] and [30], and the auxiliary package for learning procedures caret [31] have been used to implement the procedures.

Programming languages

The simulation experiment has been developed combining the programming languages Python and R, which are probably the most used programming languages in data science. Python is used for data acquisition and transformation, while R fits the FMM models. Several solutions have been studied for the coupling between them that could, at the same time, be computationally effective, robust, and simple. A basic outline on the matter is presented in [32]. While certain libraries provide tools for the coupling of the two languages, such as [33], [34] and [35], these solutions are not sufficiently refined, and bash scripting was finally used.

Results

Simulation experiment design

In the first stage, Python simulates AP signals using a modified HH model implementation based on the one available in the Neurodynex package [21]. The original implementation offers several features, such as a detailed evolution of the voltage-gated variables m, n, h , or the application of input stimuli with different amplitude and shape. A modification was implemented to facilitate changes in the model parameters. The analyzed signal spans 60 ms. A short square input stimulus I , of just 1 ms, has been applied in the tenth ms of the simulation. See [36] to learn about the most used stimulus types. Five values for the strength of the stimulus have been selected, $I = \{0, 4.5, 7, 9.5, 12\} \mu A$, leading to a total of 5000 experiments. Different configurations of the most influential parameters have been considered according to a factorial experiment design. In every experiment, each of these parameters takes a random value from a set of preselected values. These values are experimentally observed in nature, as described in [7] and detailed in Table 1.

Table 1. Parameters varied in the Hodgkin-Huxley simulations according to the designed factorial experiment design.

Structural Parameters	Kinetic Parameters
$C = 1 \mu F/cm^2$	$\tilde{a}_n = \{0.85, 0.95, 1.05, 1.15\}$
$V_{Na} = 50 \text{ mV}$	$\tilde{b}_n = \{0.7, 0.85, 1, 1.15, 1.3\}$
$V_K = -77 \text{ mV}$	$\tilde{a}_m = \{0.8, 0.9, 1, 1.1, 1.2, 1.3\}$
$V_L = -54 \text{ mV}$	$\tilde{b}_m = \{0.7, 0.85, 1, 1.15, 1.3\}$
$g_{Na} = \{64, 92, 120, 148, 176, 204, 232, 260\} ms/cm^2$	$\tilde{a}_h = 1$
$g_K = \{27, 30, 33, 36, 39, 42, 45, 48\} ms/cm^2$	$\tilde{b}_h = 1$
$g_L = \{0.12, 0.215, 0.31, 0.405, 0.5\} ms/cm^2$	

In a second stage, using R, the signals are preprocessed to assure that complete APs are analyzed. We proceed as follows: the first AP is discarded if it occurs before the input stimulus has been applied; resting potential values are considered instead of the discarded data to get segments of equal length for all the trials. An analogous procedure is applied if the last AP is observed in the last 6 ms of the experiment. In the exceptional case, where only a pre-stimulus AP is observed, the original signal is analyzed. It is relevant to note that discarding the first and/or last AP in the experiments is not a limitation of the approach, as we are assuming equally-shaped APs.

In a third step, an FMM_{ST} model is fitted to the data using the package [23].

Simulation experiment results

3613 out of the total 5000 simulated HH experiments have been analyzed using the FMM approach as they have at least one AP. In Fig. 2, representative signals with one to four spikes are plotted along with the FMM_{ST} model predictions. There is a relevant Dominant Component that represents the depolarization and repolarization; and a second component that accounts for the hyperpolarization. A summary of the main statistics and parameter estimates are given in Table 2. The values in the table show the high prediction accuracy. In particular, the $R^2_{\text{FMM}_{\text{ST}}}$ global mean (standard deviation) is equal to 0.8438 (0.0683), and the $R^2_{\text{FMM}_{\text{ST}}}$ is equal to 0.9864 (0.0054). These values quantify what Fig. 2 shows. Table 2 also shows interesting differences in parameter configurations between signals with a different number of APs.

Table 2. Means and standard deviations for R^2 values and parameter estimators.

	Number of APs in the Signal						ALL
	1	2	3	4	5	6	
$R^2_{\text{FMM}_{\text{ST}}}$	0.817 (0.061)	0.817 (0.089)	0.857 (0.056)	0.881 (0.028)	0.886 (0.027)	0.902 (0.030)	0.852 (0.058)
$R^2_{\text{FMM}_{\text{ST}}}$	0.985 (0.005)	0.986 (0.009)	0.989 (0.006)	0.990 (0.003)	0.989 (0.003)	0.988 (0.004)	0.988 (0.005)
M	39.211 (3.936)	80.448 (22.861)	96.346 (24.920)	117.824 (23.591)	156.227 (22.320)	183.459 (19.165)	89.650 (45.465)
A_m^A	53.342 (4.258)	50.831 (4.215)	50.541 (4.445)	52.115 (3.646)	54.033 (3.048)	54.313 (2.629)	52.397 (4.182)
A_m^B	16.0133 (5.954)	25.292 (5.106)	24.396 (4.077)	22.549 (3.636)	20.986 (3.382)	18.169 (2.987)	20.538 (5.826)
β^A	2.780 (0.279)	2.529 (0.214)	2.560 (0.180)	2.582 (0.152)	2.582 (0.131)	2.589 (0.152)	2.645 (0.232)
$\cos(\beta^B)$	0.606 (0.347)	-0.142 (0.297)	-0.110 (0.232)	-0.059 (0.203)	-0.040 (0.224)	-0.029 (0.243)	0.162 (0.424)
$\sin(\beta^B)$	-0.573 (0.428)	-0.941 (0.092)	-0.960 (0.111)	-0.977 (0.028)	-0.973 (0.037)	-0.970 (0.049)	-0.831 (0.320)
ω^A	0.041 (0.005)	0.030 (0.006)	0.029 (0.005)	0.030 (0.004)	0.030 (0.004)	0.030 (0.005)	0.033 (0.007)
ω^B	0.136 (0.066)	0.086 (0.044)	0.070 (0.020)	0.066 (0.012)	0.0689 (0.012)	0.076 (0.016)	0.092 (0.052)
d_m^{AP}		1.515 (0.767)	1.488 (0.335)	1.125 (0.126)	0.814 (0.068)	0.636 (0.021)	0.784 (0.643)
d_m^{AB}	0.039 (0.023)	0.034 (0.071)	0.029 (0.012)	0.032 (0.011)	0.036 (0.014)	0.041 (0.016)	0.034 (0.020)

Fig 2. Neuronal signals simulated with the HH model and the estimated FMM_{ST} signals in red (left). Components of the fitted FMM_{ST} (right).

In particular, to illustrate how parameter configuration differentiates between excitable and oscillatory neurons in comparison with the HH parameters, two radar-plots corresponding to the HH and FMM models' main parameters are represented in Fig. 3. The HH model parameters discriminate the two neuron types by biochemical properties. Specifically, the oscillatory experiments have primordially higher values in g_{Na} , $\tilde{b}_m$ and $\tilde{a}_n$. However, the FMM parameters differentiate the two types by the shape of the APs. It seems that excitable signals have both the components with a higher kurtosis and skewness ($\omega^A, \omega^B, \beta^A, \beta^B$) and a B component with a smaller amplitude (A_m^B).

Fig 3. Median values of the HH parameters and FMM_{ST} parameters (defined in Table 3) by neuron type. The represented interval for each parameter is between the 10% and 90% percentiles.

Next, the analysis related to the $(\mathcal{S}, \mathcal{K})$ parameters is presented.

On the one hand, scatter plots of $\mathcal{K}$ against $\mathcal{S}$ across intensities have been performed, imitating those in [7], in Fig. 4. Similar comments to those in the cited paper arise. Moreover, these plots also illustrate how the number of excitable and oscillatory experiments increase with the intensity. On the other hand, the potential of the model to predict the pair $(\mathcal{S}, \mathcal{K})$ is shown. Two different sets of predictors: τ^A and $\tau^A + \tau^B$, defined in Table 3 and different Supervised Learning procedures have been considered. Note that the LR and SVM approaches assume that the predictors are euclidean, but β is a circular parameter. Then, $\cos(\beta^B)$ and $\sin(\beta^B)$ are considered instead of β^B . Furthermore, β^A is considered as euclidean as it takes values concentrated in a small arc. Other predictors, derived from those in $\tau^A + \tau^B$, have been considered in the preliminary analysis but were eventually discarded using the principle of parsimony as only the results of LR improved lightly.

Variable selection methods have been used, a stepwise AIC for LR and caret's RFE [31] for the rest. While in the former case, some variables are discarded, all the variables are included in the other models.

Fig 4. (S, K) values by stimulus amplitude and neuron type.

Table 3. Feature sets used in the prediction of $\mathcal{S}$ and $\mathcal{K}$.

τ^A		τ^B	
M	Baseline level of the signal.		
A_m^A	Median of the amplitude A_S^A , $S = 1, \dots, s$.	A_m^B	Median of the amplitude A_S^B , $S = 1, \dots, s$.
β^A	Skewness of the A waves.	β^{B*}	Skewness of the B waves.
ω^A	Kurtosis of the A waves.	ω^B	Kurtosis of the B waves.
d_m^{AP}	Median of d_S^{AP} , $S = 1, \dots, s - 1$.	d_m^{AB}	Median of d_S^{AB} , $S = 1, \dots, s$.
s	Number of APs in the signal.		

*: In linear procedures, $\cos(\beta^B)$ and $\sin(\beta^B)$ are used.

Furthermore, ten-fold cross-validation is considered for comparative and validation purposes. The dataset is divided into ten equally sized splits. In ten iterations, nine of the subsets are used to train the model, while the tenth serves as a test as in [25] and [37]. In addition, the Generalized Degrees of Freedom (GDFs) of the models have been calculated, as proposed in [38], to measure the underlying complexity.

Table 4 provides a summary of the results for the $\mathcal{S}$ and $\mathcal{K}$ models. It can be seen that the parameters associated with the second component, τ^B , significantly increase the prediction accuracy compared with that obtained using only parameters associated to component A . Regarding the different procedures, at one end, the relatively bad results for LR provides evidence that the relation between the predictors and $(\mathcal{S}, \mathcal{K})$ is not linear. At the other end, SVM gives the most accurate prediction, with more than 95% and 94% of explained variance for $\mathcal{S}$ and $\mathcal{K}$ respectively. The RF and GBM results are between those for LR and SVM; while RF is comparable to GBM in terms of interpretability and complexity, the attained accuracy is less. However, although GBM procedures remain more complex and slightly less accurate than SVM, the predictions are interpretable. Specifically, the most relevant predictors to explain $\mathcal{S}$ are β^A and ω^A , which implies that $\mathcal{S}$ is well predicted by the AP shape. In the case of $\mathcal{K}$, the most relevant predictors are M , d_m^{AP} and s .

Table 4. R^2 and GDFs for the LR, RF, SVM and GBM models to predict $\mathcal{S}$ and $\mathcal{K}$.

			LR	RF	SVM	GBM
$\mathcal{S}$	Predictors: τ^A	R^2	0.6825	0.8421	0.8672	0.8576
		GDF	7.05	2053.13	301.28	1362.50
	Predictors: $\tau^A + \tau^B$	R^2	0.7933	0.9127	0.9548	0.9467
		GDF	12.21	2123.83	242.48	2699.63
			LR	RF	SVM	GBM
$\mathcal{K}$	Predictors: τ^A	R^2	0.6542	0.8152	0.8446	0.8309
		GDF	6.74	2022.95	297.10	1766.07
	Predictors: $\tau^A + \tau^B$	R^2	0.7548	0.8877	0.9413	0.9350
		GDF	11.11	2113.74	381.73	2845.45

Predictive analysis for other HH parameters has also been performed. The results for sodium parameters are included in Table 5. Sodium parameters have been selected as they are more accurately predicted and are the engine of the neuron excitability, as some authors such as [39] and [8] claim. It is interesting to note that the $\tilde{a}_m$ and $\tilde{b}_m$ are directly related to the first and second FMM components, respectively, while g_{Na} is

related to both. The models' accuracy is not as high as for (S, K) , which are more stable parameters.

Table 5. R^2 values for LR, RF, SVM and GBM models to predict sodium parameters.

		LR	RF	SVM	GBM
g_{Na}	Predictors: τ^A	0.5916	0.7207	0.7403	0.7309
	Predictors: $\tau^A + \tau^B$	0.7106	0.7943	0.8276	0.8147
		LR	RF	SVM	GBM
$\tilde{a}_m$	Predictors: τ^A	0.3606	0.5801	0.6011	0.5909
	Predictors: $\tau^A + \tau^B$	0.3760	0.7115	0.7810	0.7538
		LR	RF	SVM	GBM
$\tilde{b}_m$	Predictors: τ^A	0.2105	0.4013	0.3593	0.3901
	Predictors: $\tau^A + \tau^B$	0.6521	0.8374	0.9220	0.9103

Real Data

The SGAMP database, firstly used in [20] and publicly accessible at [40], contains single-unit neuronal recordings of North Atlantic squid (*Loligo pealei*) giant axons in response to stimulus currents. The database has been extensively used in works. Among the recent ones are [41] and [42]. The experiments where the applied stimulus is a short square stimulus, the same as in the HH experimentation, have been selected for the analysis. Five signals have been extracted from 4 trials of 3 different axons the stimulus amplitude being equal to $I = 5 \mu A$ in the five cases. The length of the analyzed segment is also equal to that of the simulated HH signals, 60 ms, to facilitate comparisons.

Four different FMM models have been fitted to the signals: an $FMM_{S^{**}}$, $FMM_{S^{*}}$, FMM_{ST} and an $FMM_{ST^{*}}$. Moreover, it is assumed that $A^A - A^B < C$, where C is 0.70 times the maximum difference obtained in previous iterations of the algorithm. For comparative proposes, Fourier models with the same number of free parameters as the FMM models, denoted as FD_a where a is the number of harmonics, have also been fitted.

Fig. 5 shows the $FMM_{ST^{*}}$ and the corresponding Fourier predictions for a representative signal. The R^2 values and the number of free parameters are given in Table 6, for the five signals and eight models. The figures in the table show that the Dominant Component explains much of the variability here, much more than it does in the HH experimentation. Furthermore, the models with restrictions in the A parameters are as accurate as the others without them. As such, the former models, being the simpler ones, are preferred due to the parsimony principle. Besides, the FD_a models make much less accurate predictions. Moreover, Table 7 gives the parameter values for the $FMM_{ST^{*}}$ model of the five signals. Compared to the values in table 2, in particular to those corresponding to signals with 4 APs, some interesting differences can be highlighted, in particular in A^A , A^B , β^A , and β^B .

Finally, a principal component analysis has been performed using the basic set of FMM parameters from the HH experiments. A plane for the first two principal

Table 6. R^2 values of the different FMM and FD models for the signals extracted from the SGAMP database.

	N ^o of parameters	Experiment IDs				
		a1t18_1	a1t18_2	a1t22	a3t03	a4t13
FMM_{S**}	8	0.9435	0.9457	0.9472	0.9512	0.9324
FMM_{S*}	11	0.9485	0.9511	0.9476	0.9566	0.9429
FMM_{ST*}	15	0.9862	0.9848	0.9880	0.9936	0.9910
FMM_{ST}	21	0.9865	0.9852	0.9882	0.9937	0.9910
FD₄	9	0.3635	0.3609	0.3587	0.4080	0.2809
FD₅	11	0.3648	0.3622	0.3594	0.4141	0.2865
FD₇	15	0.3687	0.3636	0.3614	0.4305	0.3056
FD₁₀	21	0.5676	0.5809	0.5586	0.6099	0.4695

Fig 5. A neuronal AP from the SGAMP database (axon 1, trial 18, second short square stimulus) along with the fitted signals using an FMM_{ST*} signal (red) and an FD₇ (blue). The components of the fitted FMM_{ST*} model are illustrated at the bottom.

components is given in Fig. 6, where each point represents an HH simulated signal. In that plot, the values corresponding to SGAMP data have also been plotted as they are located far from the main cloud of HH points.

299
300
301

Table 7. Parameter estimators of the FMM_{ST*} models for the SGAMP signals.

	Experiment IDs					Mean
	a1t18_1	a1t18_2	a1t22	a3t03	a4t13	
M	173.617	179.144	115.622	158.667	184.105	162.231
A^A	44.955	44.472	37.352	43.089	46.284	43.230
A^B	6.893	6.770	13.989	6.478	27.524	12.331
β^A	3.426	3.311	3.933	3.572	3.634	3.575
$\cos(\beta^B)$	-0.368	-0.387	-0.214	-0.428	-0.221	-0.324
$\sin(\beta^B)$	-0.930	-0.922	0.977	-0.903	0.975	-0.161
ω^A	0.028	0.028	0.030	0.027	0.027	0.028
ω^B	0.101	0.138	0.008	0.097	0.005	0.070
d_m^{AP}	1.016	0.985	0.987	1.033	1.049	1.014
d_m^{AB}	0.021	0.020	0.002	0.020	0.002	0.013

This latter fact, and the results shown above the percentage of variance explained by the Dominant Component, indicate that the model underlying SGAMP signals is not an HH but a simpler one, such as FitzHugh-Nagumo (see [9]).

Fig 6. Plot of the first two principal components of $\tau^A + \tau^B$ of the HH experiments. Experiments with 4 APs observed have been highlighted. The projections of the SGAMP experiments have been added to the plot with the corresponding label.

Discussion

In this paper, the FMM_{ST} model has been presented, and its potential to predict simulated and real AP signals has been proved. Our analysis shows that some differences in shape exist between HH and SGAMP signals, even though it has often been assumed that the HH model is adequate to represent the SGAMP data in the literature. Therefore, a simpler model may be more suitable to represent SGAMP signals. Moreover, the accurate prediction of HH and SGAMP signals also provides evidences that other neuronal dynamics models could also be represented using the FMM approach. Differences between models can be articulated through the differences in the FMM parameter configurations.

Furthermore, the paper shows the potential of the model to describe such neuronal features as $(S, \mathcal{K})$. In particular, the flexibility of the FMM parameters to describe different AP waveforms would contribute to solving problems where the AP characterization is needed. Specifically, it is relevant in Spike Sorting, which is the process in which spikes are grouped into clusters based on the similarity of their shapes. The resulting clusters are assumed to correspond to different neurons. Therefore, Spike Sorting is the first step in many data analyses, being one of the most critical data analysis problems in neurophysiology and has received a lot of attention in the literature. Some interesting references are [24], [43] and [44], among others.

The proposed approach is also flexible concerning the segment length. The signal can be analyzed, either for individual AP segments, which implies spike cutting as in other models, such as [13], or by analyzing a more extended length of signal, most likely with several APs, as is done in this work.

Two different lines of work could be defined for the future. On the one hand, from a theoretical perspective, a first question to solve is the implementation of restrictions with the form $d_1^{AB} = d_S^{AB}$; $S = 2, \dots, s$. They are of interest to analyze signals with equal distances between components in the APs, because reducing the number of parameters to be estimated is important when large or noisy Spike Trains are analyzed. On the other hand, from an applied perspective, many other AP real signals must be analyzed, and the questions of spike classification and clustering may be addressed. Particularly, it would be of interest to consider the FMM parameters in the problem of neuronal type classification, as they may be useful to characterize the electrophysiological mechanisms, morphological features, and genetic profiles. Some works in this line are [2], [13] and [45]. The FMM approach's potential is difficult to calibrate as many aspects remain to be researched and exploited.

Finally, a limitation of our study is that the input stimulus's timing and shape have been fixed. The influence of the stimulus type could be analyzed using the FMM approach. The α parameters are related to the firing times, and the shape parameters could be useful to detect circumstances where the shape of the APs is independent of the stimulus, as [4] suggests, and circumstances where changes happen, such as the observation of incomplete spikes. However, the question is tricky and deserves further research.

References

1. Mensi S, Naud R, Pozzorini C, Avermann M, Petersen CCH, Gerstner W. Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of neurophysiology*. 2012;107(6):1756–1775.

2. Zeng H, Sanes J. Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nature Reviews Neuroscience*. 2017;18. doi:10.1038/nrn.2017.85.
3. Trainito C, von Nicolai C, Miller EK, Siegel M. Extracellular spike waveform dissociates four functionally distinct cell classes in primate cortex. *Current Biology*. 2019;29(18):2973–2982.
4. Raghavan M, Fee D, Barkhaus PE. Generation and propagation of the action potential. In: *Handbook of clinical neurology*. vol. 160. Elsevier; 2019. p. 3–22.
5. Ori H, Hazan H, Marder E, Marom S. Dynamic clamp constructed phase diagram for the Hodgkin and Huxley model of excitability. *Proceedings of the National Academy of Sciences*. 2020;117(7):3575–3582. doi:10.1073/pnas.1916514117.
6. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*. 1952;117(4):500–544.
7. Ori H, Marder E, Marom S. Cellular function given parametric variation in the Hodgkin and Huxley model of excitability. *Proceedings of the National Academy of Sciences*. 2018;115(35):E8211–E8218. doi:10.1073/pnas.1808552115.
8. Marom S. Emergence and maintenance of excitability: kinetics over structure. *Current Opinion in Neurobiology*. 2016;40:66 – 71. doi:https://doi.org/10.1016/j.conb.2016.06.013.
9. Fitzhugh R. Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical journal*. 1961;1 6:445–66.
10. Abbott LF, Kepler TB. Model neurons: From Hodgkin-Huxley to Hopfield. In: *Statistical Mechanics of Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1990. p. 5–18.
11. Izhikevich EM. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*. 2003;14(6):1569–1572. doi:10.1109/TNN.2003.820440.
12. Ermentrout B, Terman D. *The Mathematical Foundations of Neuroscience*. 1st ed. Springer; 2010.
13. Teeter C, Iyer R, Menon V, Gouwens N, Feng D, Berg J, et al. Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*. 2018;9(1):1–15.
14. Brunton BW, Beyeler M. Data-driven models in human neuroscience and neuroengineering. *Current Opinion in Neurobiology*. 2019;58:21 – 29. doi:https://doi.org/10.1016/j.conb.2019.06.008.
15. Picinbono B. On instantaneous amplitude and phase of signals. *IEEE Transactions on Signal Processing*. 1997;45(3):552–560. doi:10.1109/78.558469.
16. Boashash B. *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*. 1st ed. Academic Press; 2003.
17. Rueda C, Larriba Y, Peddada SD. Frequency Modulated Möbius Model Accurately Predicts Rhythmic Signals in Biological and Physical Sciences. *Scientific Reports*. 2019;9(1):1–10.

18. Rueda C, Rodríguez-Collado A, Larriba Y. A novel wave decomposition for oscillatory signals; 2020. Available from: <https://arxiv.org/abs/2006.15974>.
19. Rueda C, Larriba Y, Lamela A. The hidden wave in the ECG uncovered: a sound automated interpretation method; 2020. Available from: <https://arxiv.org/abs/2005.10173>.
20. Paydarfar D, Forger DB, Clay JR. Noisy Inputs and the Induction of On-Off Switching Behavior in a Neuronal Pacemaker. *Journal of Neurophysiology*. 2006;96(6):3338–3348. doi:10.1152/jn.00486.2006.
21. Gerstner W, Kistler WM, Naud R, Paninski L. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. New York, NY, USA: Cambridge University Press; 2014.
22. Estumano D, Orlande HR, Colaço M, Ritto T, Diaz J, Dulikravich G. Bayesian Estimation of Parameters in Hodgkin-Huxley's Model of Biomedical Electric Signals. In: 2nd International Symposium on Uncertainty Quantification and Stochastic Modeling; 2014.
23. Lamela A. FMM: Rhythmic patterns modeling by FMM; 2020. Available from: <https://CRAN.R-project.org/package=Fmm>.
24. Rey HG, Pedreira C, Quiroga R. Past, present and future of spike sorting techniques. *Brain research bulletin*. 2015;119:106–117.
25. Hastie T, Tibshirani R, Jerome F. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. Springer; 2009.
26. Izenman AJ. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. 1st ed. Springer Publishing Company, Incorporated; 2008.
27. R Core Team. *R: A Language and Environment for Statistical Computing*; 2013. Available from: <http://www.R-project.org/>.
28. Breiman L, Cutler A, Liaw A, Wiener M. Package 'randomForest' Manual. CRAN; 2018.
29. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*. 2004;11(9):1–20.
30. Ridgeway G. *Generalized Boosted Models: A Guide to the GBM Package*. Compute. 2005;1:1–12.
31. Kuhn M. Package 'caret' Manual. New London: CRAN; 2018.
32. Pandey P. From 'R vs Python' to 'R and Python'. *Towards Data Science*; 2019. Available from: <https://towardsdatascience.com/from-r-vs-python-to-r-and-python-aa25db33ce17>.
33. Gautier L. rpy2 - R in Python; 2019. Available from: <https://rpy2.bitbucket.io/>.
34. Curtis V. RWinOut; 2017. Available from: <https://github.com/vitorcurtis/RWinOut>.
35. Xia XQ, McClelland M, Wang Y. Pyper, A Python Package for Using R in Python. *Journal of Statistical Software, Code Snippets*. 2010;35(2):1–8.

36. Allen Brain Institute. Allen Cell Types Database - Electrophysiology. Technical report; 2015. Available from: http://help.brain-map.org/download/attachments/8323525/CellTypes_Ephys_Overview.pdf?version=2&modificationDate=1508180425883&api=v2.
37. Kubat M. An Introduction to Machine Learning. 1st ed. Springer Publishing Company, Incorporated; 2015.
38. Ye J. On Measuring and Correcting the Effects of Data Mining and Model Selection. *Journal of The American Statistical Association*. 1998;93:120–131. doi:10.1080/01621459.1998.10474094.
39. Silva J. In: Ruben PC, editor. Slow Inactivation of Na⁺ Channels. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 33–49. Available from: https://doi.org/10.1007/978-3-642-41588-3_3.
40. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*. 2000 (June 13);101(23):e215–e220.
41. Forger D, Paydarfar D, Clay J. Optimal Stimulus Shapes for Neuronal Excitation. *PLoS computational biology*. 2011;7:e1002089. doi:10.1371/journal.pcbi.1002089.
42. Clay J, Forger D, Paydarfar D. Ionic Mechanism Underlying Optimal Stimuli for Neuronal Excitation: Role of Na⁺ Channel Inactivation. *PloS one*. 2012;7:e45983. doi:10.1371/journal.pone.0045983.
43. Caro-Martín CR, Delgado-García JM, Gruart A, Sánchez-Campusano R. Spike sorting based on shape, phase, and distribution features, and K-TOPS clustering with validity and error indices. *Scientific reports*. 2018;8(1):1–28.
44. Souza BC, Lopes-dos Santos V, Bacelo J, Tort AB. Spike sorting with Gaussian mixture models. *Scientific reports*. 2019;9(1):1–14.
45. Ghaderi P, Marateb H, Safari MS. Electrophysiological Profiling of Neocortical Neural Subtypes: A Semi-Supervised Method Applied to in vivo Whole-Cell Patch-Clamp Data. *Frontiers in Neuroscience*. 2018;12. doi:10.3389/fnins.2018.00823.