Published in final edited form as:

Int J Neural Syst. 2015 March ; 25(2): 1550001. doi:10.1142/S012906571550001X.

Principal Dynamic Mode Analysis of the Hodgkin–Huxley Equations

Steffen E. Eikenberry^{*} and Vasilis Z. Marmarelis[†]

Department of Biomedical Engineering University of Southern California 1042 Downey Way, Los Angeles, CA 90089, USA

Abstract

We develop an autoregressive model framework based on the concept of Principal Dynamic Modes (PDMs) for the process of action potential (AP) generation in the excitable neuronal membrane described by the Hodgkin–Huxley (H–H) equations. The model's exogenous input is injected current, and whenever the membrane potential output exceeds a specified threshold, it is fed back as a second input. The PDMs are estimated from the previously developed Nonlinear Autoregressive Volterra (NARV) model, and represent an efficient functional basis for Volterra kernel expansion. The PDM-based model admits a modular representation, consisting of the forward and feedback PDM bases as linear filterbanks for the exogenous and autoregressive inputs, respectively, whose outputs are then fed to a static nonlinearity composed of polynomials operating on the PDM outputs and cross-terms of pair-products of PDM outputs. A two-step procedure for model reduction is performed: first, influential subsets of the forward and feedback PDM bases are identified and selected as the reduced PDM bases. Second, the terms of the static nonlinearity are pruned. The first step reduces model complexity from a total of 65 coefficients to 27, while the second further reduces the model coefficients to only eight. It is demonstrated that the performance cost of model reduction in terms of out-of-sample prediction accuracy is minimal. Unlike the full model, the eight coefficient pruned model can be easily visualized to reveal the essential system components, and thus the data-derived PDM model can yield insight into the underlying system structure and function.

Keywords

Volterra model; nonlinear modeling; neuronal modeling; Laguerre expansions; autoregressive model; nonparametric model; basis expansion; bootstrap

1. Introduction

The Hodgkin–Huxley (H–H) equations¹ describe the process of action potential (AP) generation in the excitable neuronal membrane in response to a current injection, which is dominated by highly nonlinear ion channel dynamics. The H–H equations have been widely studied since their proposal over 50 years ago, and remain the canonical mathematical description of AP generation.

^{*}seikenbe@usc.edu. [†]Corresponding author. vzm@usc.edu.

This paper follows our previous work² in which we developed the Nonlinear Auto-Regressive Volterra-type (NARV) model for neuronal membrane dynamics, which we applied to the H–H equations. The NARV model is a nonlinear “black-box” description of the input–output relation between injected current and membrane potential for a spiking neuron. Mathematically, it is based on the modified Volterra series and the input–output relation is fully defined by its “Volterra kernels”. The Volterra series³ is a functional power series expansion that relates any input function, $x(t)$, $t \in \mathbb{R}$, to an output scalar, $y(t)$, and this relation can be expressed in hierarchical convolutional form, where each hierarchy represents an order of nonlinearity (see Sec. 2).

The venerable theory of functionals was largely developed by the Italian mathematician Vito Volterra, who first introduced the basic idea of a “function of a line” in 1887.⁴ This idea was further extended and abstracted in the following decades; notable works include a 1910 paper⁵ and a series of lectures published in 1913 (see the preface to the English edition in Ref. 3). A more thorough historical overview, as well as the most complete and accessible treatment of the subject is given in Volterra's 1930 book,³ based on a series of lectures delivered in 1925. This abstract concept has proven to have great practical utility, and the Volterra series is applicable to a wide variety of physiologic systems in which input is transformed into some output. An extensive, but still partial, review of applications may be found in Ref. 6.

The concept of principal dynamic modes (PDMs) posits that the (time-dependent) dynamics of a system may be represented by a small set of basis functions, unique to that system.⁷ These PDMs may be estimated from the NARV model and used as a basis for a PDM model that is mathematically equivalent in structure to the NARV model, but having the crucial advantage of an equivalent modular representation that can reveal the functional characteristics of the system. In this paper, we apply the PDM concept to the H–H system to demonstrate that it allows very significant reduction in model complexity at minimal cost to performance, while much better revealing the underlying system's structural dynamical relationships among the exogenous and autoregressive components. Ultimately, the second-order NARV model, with 65 free coefficients, can be reduced to an eight parameter PDM model with only a marginal performance cost, or to only four parameters at somewhat higher cost. Note that this also compares favorably to the 24-parameter system of differential equations that defines the H–H system. The proposed PDM methodology is intended as a general method which may be used to describe any spiking neuron, simulated or real, and the H–H system is used as a well-studied test case.

Numerous Volterra-style models have been proposed to describe the input–output relationship of injected current ($\mu\text{A cm}^{-2}$) to membrane potential (mV) for both the H–H equations and real neurons^{8–19}; these references are briefly reviewed in Ref. 2. While Volterra models can in principle describe any finite-memory input–output system, the essentially autonomous nature of the AP makes the direct application of the Volterra series problematic. We have previously argued that the subthreshold and suprathreshold dynamics of the H–H equations represent distinct dynamical regimes. In the subthreshold regime, the output is a function of the recent exogenous input, and the standard Volterra series applies well. However, upon AP initiation the underlying ion channel dynamics result in a highly

stereotyped waveform and the output becomes effectively decoupled from the input. To compensate for this, we proposed the NARV model, in which the output, $y(n)$, is thresholded as

$$\hat{y} = y(t) H[y(t) - \theta], \quad (1)$$

where $H[\cdot]$ is the Heaviside step function, and this thresholded output, $\hat{y}$, is fed back as a second input to a two-input Volterra model. The NARV model is estimated by the Laguerre expansion technique (LET),^{7,20} whereby the kernels are expanded on a basis of discrete Laguerre functions (DLFs),²¹ dramatically reducing the dimensionality of the problem. This also yields an equivalent model architecture in which the exogenous and (thresholded) autoregressive inputs are passed through linear filterbanks consisting of the Laguerre bases, and the outputs of these filterbanks are passed to a static-nonlinearity which gives the final output (see Fig. 2 of Ref. 2).

While displaying excellent predictive performance, the NARV model is difficult to interpret, either by examining the reconstructed Volterra kernels, or through its modular representation, which consists of 10 DLFs and 65 coefficients. This motivates the application of the PDM concept, in which we estimate a functional basis for kernel expansion that is specific to the particular system under study. We separately estimate forward and feedback PDM bases from the forward and feedback Volterra kernels, using singular value decomposition. The two PDM bases then take the place of the corresponding Laguerre bases in the forward and feedback filterbanks. The nonlinearity is further decomposed conceptually into forward and feedback self- and cross-terms. Thus, we move from a generic modular representation to a system-specific modular representation, but preserve the basic mathematical framework.

The PDM model is also highly amenable to reduction, because we may select a highly influential subset of PDMs as a basis. The PDMs can be ranked by singular value, which gives a rough indication of their relative importance, and the subset may be selected using this criterion.²² However, at least in this work, we have found this ranking to be unreliable (with the exception of the highest ranked PDM), and we use a leave-one-out/add-one-in (LOO/AOI) algorithm to identify the reduced PDM basis set. This algorithm identifies two forward and four feedback PDMs that are significant, yielding a 27-parameter (27-P) basis-reduced model with performance comparable to that of the NARV model.

Given the reduced PDM bases, we perform further model reduction by pruning the coefficients of the static nonlinearity. We iterate through each coefficient and leave it out of the full model; if the omission of a coefficient results in a statistically significant reduction in performance, it is retained, otherwise it is omitted. Several summary statistics are used to quantify model performance, and bootstrap procedure is used to estimate the variance of these statistics. Using these statistics and a bootstrap procedure, we thoroughly demonstrate that the basis-reduced and coefficient-pruned PDM models give performance comparable to or only minimally inferior to that of the full model. The principal focus of this paper is on four particular models: the 65-parameter (65-P) full PDM model, the 27-P basis-reduced model, and 8-P and 4-P coefficient-pruned models.

We have compared final coefficient-pruned models where the PDM basis reduction step is performed first versus using the entire PDM basis and find that reducing the PDM basis set first always results in a better final model. Thus, eliminating coefficients associated with insignificant PDMs before pruning appears to improve the performance of the pruning algorithm by pre-eliminating these spurious coefficients.

We have also generated curves plotting the number of retained coefficients versus the performance metrics, where those coefficients found to have the least influence on predictive ability are removed in succession. We find that the inclusion of a significant number of coefficients actually reduces model performance, probably due to overfitting.

Examining the structure of the reduced and pruned PDM-based models shows that, within the subthreshold regime, the H–H membrane acts as a “leaky integrator” with a memory of approximately 5 ms, in accordance with the widely posited leaky integrating characteristic of the H–H membrane.²³ The leaky integrating characteristic inferred from the data-based analysis, however, is not a simple exponential and appears to have three different time constants. Furthermore, unlike leaky-integrator neurons, the PDM model does not simply reset the membrane potential upon threshold crossing, but accounts for the AP and associated afterpotential and refractory period through a separable early AP waveform component, representing the invariant AP response to threshold crossing, and bilinear interactions among forward and feedback PDMs to give complex peri-AP phenomena, e.g. refractoriness. The PDM model also differs from the previously proposed Spike Response Model (SRM),¹⁵ which may be viewed as a generalized integrate-and-fire model.^{18,24} The SRM convolves the injected input with a linear integrating kernel, while the AP and feedback is accounted for by a linear AP and afterpotential kernel. The SRM lacks the bilinear cross-terms of the PDM model that have proven, in our model framework, essential to accurately representing the AP waveform and accurate model predictions.

We have also generated H–H data-sets with the underlying H–H ion channel dynamics altered, specifically the maximum sodium and potassium conductances. We impose a minimal four coefficient architecture, using the “universal PDM basis” identified on the standard H–H system, and examine how the coefficients change systematically with changes in the ionic conductances. We find a clear pattern in which increases in sodium conductance diminish coefficients associated with the AP and refractoriness, whereas increased potassium conductance has the opposite effect. The overall effects, however, are not simple inverses of each other. Thus, changes in the underlying biophysics are reflected in an orderly fashion in the highly reduced data-based PDM model.

The proposed methodology represents a general method for nonlinear system identification that may, at least in principle, be applied to systems with multiple inputs and outputs, and the outputs may or may not be autoregressive. The method is practically useful in that the same mathematical framework is used at all steps of model identification. That is, the input–output relation is expressed as a family of functional bases that transform input as linear filterbanks, and an associated polynomial nonlinearity. This polynomial may be conceptually subdivided into various self- and feedback terms, allowing better visualization of the model structure.

In sum, we propose a methodology for system identification based on the Volterra series and PDM concepts that (1) uses the same basic mathematical framework as the NARV model and modified discrete Volterra (MDV) model that has been widely used in physiological modeling,⁶ (2) is amenable to dramatic reduction in model complexity, as defined by the number of free model coefficients, by first reducing the PDM basis sets and then pruning the coefficients of the static nonlinearity, (3) has a modular structure specific to the system under study that can be visualized and interpreted much more readily than kernel-based Volterra models, and (4) uses a flexible and computationally straightforward bootstrap procedure for statistical evaluation of all model components (PDMs and expansion coefficients). Application of the method to the H–H system yields an eight-parameter PDM model that is comparable to the full 65-parameter NARV model.

2. Methods

2.1. Data preparation

As in Ref. 2, we have generated time-series data for the H–H model¹ with injected current, $I(t)$, as input, and membrane potential, $V(t)$, as output. The H–H model describes how ionic fluxes generated by voltage-gated potassium and sodium channels and a so-called “leak” channel give rise to changes in membrane potential. Briefly, the model considers the potassium gate to be composed of four subunits (n) which are activated by membrane depolarization, while the sodium gate has three activating subunits (m) and one deactivating subunit (h), allowing the sodium current to be transiently activated by depolarization. The full model is given as:

$$C_M \frac{dV}{dt} = I(t) - \bar{g}_K n^4 (V - V_K) - \bar{g}_{Na} m^3 h (V - V_{Na}) - \bar{g}_l (V - V_l), \quad (2)$$

$$\frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n, \quad (3)$$

$$\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m, \quad (4)$$

$$\frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h, \quad (5)$$

where

$$\alpha_n = \frac{0.1 - 0.01V}{\exp(1 - 0.1V) - 1}, \quad \beta_n = 0.125 \exp\left(\frac{-V}{80}\right), \quad (6)$$

$$\alpha_m = \frac{2.5 - 0.1V}{\exp(2.5 - 0.1V) - 1}, \quad \beta_m = 0.4 \exp\left(\frac{-V}{18}\right), \quad (7)$$

$$\alpha_h = 0.07 \exp\left(\frac{-V}{20}\right), \quad \beta_h = \frac{1}{\exp(3 - 0.1V) + 1}. \quad (8)$$

The maximum potassium and sodium conductances are given by $\bar{g}_K$ and $\bar{g}_{Na}$, respectively. For a fixed membrane potential, n , m , and h each relax exponentially to their (voltage-dependent) steady-state values,

$$i_\infty = \frac{\alpha_i}{\alpha_i + \beta_i}, \quad (9)$$

with exponential time constant

$$\tau_i = \frac{1}{\alpha_i + \beta_i}, \quad (10)$$

where $i = n, m, h$. The channel time constants can be altered by multiplying α_i and β_i by the same constant factor. For the default parameter values, see either Ref. 2 or Ref. 1.

We generate data-sets using white-noise current values drawn from a normal distribution with zero-mean and standard deviation $\sigma = 32 \mu\text{A cm}^{-2}$, with new values chosen at 1 kHz. The sampling interval, T , is 0.2 ms.

We also generate several data-sets with altered H-H parameters, namely, the maximum potassium and sodium conductances, $\bar{g}_K$ and $\bar{g}_{Na}$, and the (derived) time constants for channel activation, τ_n , τ_m , and τ_h .

The training data-sets, from which the model coefficients are estimated, are 16,384 ms in length, and we generate 10 8,192 ms testing data-sets for model validation. We refer to all 10 concatenated as the *full testing data-set*.

2.2. Mathematical model

2.2.1. Basic model structure—The proposed data-derived models consider the *exogenous* injected current input, $x(n)$ ($\mu\text{A cm}^{-2}$), and the membrane potential output, $y(n)$ (mV). The thresholded output,

$$\hat{y}(n) = y(n) H[y(n) - \theta], \quad (11)$$

where $H[\cdot]$ is the Heaviside step function and θ is the threshold, is fed back as the *autoregressive* input.

The basic model framework for both the NARV model and the PDM models are mathematically equivalent in the general case, and the structure is derived from the basic kernel-based Volterra model as follows.

The kernel model consists of a two-input Volterra model, with the autoregressive component as the second input. The two-input Volterra model defines a set of self-kernels for each input

and a set of cross-kernels defining the (nonlinear) interactions between the inputs. The expansion in discrete-time to second order is:

$$\begin{aligned}
 y(n) = & k_{0,0} + T \sum_{m=0}^M k_{1,0}(m) x(n-m) \\
 & + T \sum_{r=1}^R k_{0,1}(r) \hat{y}(n-r) \\
 & + T^2 \sum_{m_1=0}^M \sum_{m_2=0}^M k_{2,0}(m_1, m_2) \\
 & \times x(n-m_1) x(n-m_2) \\
 & + T^2 \sum_{r_1=1}^R \sum_{r_2=1}^R k_{0,2}(r_1, r_2) \hat{y}(n-r_1) \hat{y}(n-r_2) \\
 & + T^2 \sum_{m=0}^M \sum_{r=1}^R k_{1,1}(m, r) x(n-m) \hat{y}(n-r),
 \end{aligned} \tag{12}$$

where M and R are the memory extents for the exogenous and autoregressive components, respectively, T is the sampling interval, $k_{0,0}$ is the zeroth-order kernel, $k_{1,0}$ and $k_{2,0}$ are the first- and second-order forward self-kernels, $k_{0,1}$ and $k_{0,2}$ are the first- and second-order feedback self-kernels, and $k_{1,1}$ is the second-order cross-kernel. The model may be extended to arbitrary order, and in Ref. 2 we examined the third-order model with third-order cross-terms in the NARV model context.

To reduce dimensionality, we choose two sets of discrete-time basis functions,

$\{b_j^{(x)}(m)\}_{j=0}^{L_x-1}$ and $\{b_l^{(y)}(r)\}_{l=0}^{L_y-1}$, corresponding to the two inputs $x(n)$ and $\hat{y}(n)$, to expand the kernels on. From substitution and the symmetry of the Volterra kernels, we obtain the polynomial expansion:

$$\begin{aligned}
 y(n) = & c_{0,0} + \sum_{j=0}^{L_x-1} c_{1,0}(j) v_j^{(x)}(n) + \sum_{l=0}^{L_y-1} c_{0,1}(l) v_l^{(y)}(n) \\
 & + \sum_{j_1=0}^{L_x-1} \sum_{j_2=0}^{j_1} c_{2,0}(j_1, j_2) v_{j_1}^{(x)}(n) v_{j_2}^{(x)}(n) \\
 & + \sum_{l_1=0}^{L_y-1} \sum_{l_2=0}^{l_1} c_{0,2}(l_1, l_2) v_{l_1}^{(y)}(n) v_{l_2}^{(y)}(n) \\
 & + \sum_{j=0}^{L_x-1} \sum_{l=0}^{L_y-1} c_{1,1}(j, l) v_j^{(x)}(n) v_l^{(y)}(n),
 \end{aligned} \tag{13}$$

where $\{v_j^{(x)}\}_{j=0}^{L_x-1}$ and $\{v_l^{(y)}\}_{l=0}^{L_y-1}$ are sets of transformed inputs obtained by convolution of $x(n)$ and $\hat{y}(n)$ with their associated bases:

$$v_j^{(x)}(n) = T \sum_{m=0}^M b_j^{(x)}(m) x(n-m), \tag{14}$$

$$v_l^{(y)}(n) = T \sum_{r=1}^R b_l^{(y)}(r) \hat{y}(n-r). \tag{15}$$

We also set $c_{0,0} = 0$, as we expect no deviation from membrane resting potential (0 by definition) in the absence of a current input. This is the general form for both the NARV and PDM models, and can be represented modularly as two linear filterbanks, consisting of the chosen bases, receiving exogenous and autoregressive inputs and whose outputs are fed to a static nonlinearity in the form of a polynomial expansion (see Fig. 2 of Ref. 2). For the present work, we only consider a second-order expansion (equivalent to a second-order Volterra model). Upon estimating the polynomial coefficients the equivalent Volterra kernels may easily be recovered. A much more detailed treatment of this derivation and the basics of Volterra series is given in Ref. 2.

Any set of basis functions may be chosen, and the fundamental difference between the NARV and PDM models is that in the former a more generic set of discrete Laguerre functions (DLFs) is used, the so-called Laguerre expansion technique,⁷ while in the latter the basis is specific to the system under study. However, note that the exponential relaxation characteristic of the DLFs with adjustable parameter α allows the Laguerre basis to be tuned to an individual system, and thus the Laguerre basis is much more versatile than a fixed generic basis. For the PDM model, the two sets of principal dynamic modes (forward and feedback) make up the basis sets, and we present the estimation of these PDMs from the NARV/kernel model in Sec. 2.2.3.

Under the PDM model, we decompose the polynomial expansion into self- and cross-terms (somewhat analogous to the self- and cross-kernels). Furthermore, the cross-terms may be forward-forward, feedback-feedback, or forward-feedback. Following Ref. 22, we also refer to these terms as the associated nonlinear functions (ANFs), when convenient. For a second-order expansion, the self-terms are quadratic, while the cross-terms are bilinear in their arguments, i.e. just a constant. An example of such a modular architecture for two forward and two feedback PDMs is given in Fig. 1. As discussed further in Sec. 2.3.2, the PDM model is also pruned by testing the significance of each expansion coefficient.

2.2.2. NARV model estimation—The first step in estimating the PDM model is to obtain an estimate of the system's Volterra kernels, which is done by estimating the NARV model using the Laguerre expansion technique. Following our previous work, when estimating the NARV model from a data-record generated with baseline H–H parameters, we fix $L_x = L_y = 5$, $\alpha_x = 0.4$, $\alpha_y = 0.7$, and $\theta = 4.5$ mV. To estimate the expansion coefficients from an input–output data record, the model is put in matrix form,

$$\mathbf{y} = \mathbf{V}\mathbf{c} + \epsilon, \quad (16)$$

where $\mathbf{y}$ is the vector of all N output samples, $\mathbf{V}$ is a matrix of transformed inputs constructed from Eq. (13) with each row corresponding to a discrete time-point, $\mathbf{c}$ is the coefficient vector, and ϵ is modeling error. The coefficient vector is then estimated by ordinary least squares or with the pseudoinverse, $\mathbf{V}^+$,

$$\hat{\mathbf{c}} = \mathbf{V}^+ \mathbf{y}. \quad (17)$$

To estimate the NARV model for data-records generated by altered ion dynamics, we also train α_x and α_y , which determine the effective memory extents of the forward and feedback model components, respectively. We have generated 2D surfaces of normalized root-mean-square-error (NRMSE) and the AP coincidence factor, Γ (see Sec. 2.4.1 for details on this performance metric), as functions of α_x and α_y and have observed that the optimal values of α_x and α_y are independent. Moreover, the NRMSE and Γ curves are generally unimodal and convex. Therefore, we can justifiably optimize the two α parameters independently using a ternary search procedure with respect to NRMSE.

2.2.3. PDM estimation—From the estimated NARV coefficients, the first- and second-order Volterra kernels are reconstructed. Several related methods for PDM estimation have been proposed.^{6,22,25} The method used here is similar to that proposed by Marmarelis *et al.*²² and is performed on both the self forward and feedback kernels to obtain a set of self and feedback PDMs as follows:

- (1) Perform eigendecomposition on the second-order kernel (this is always possible, as the second-order self-kernels are symmetric). Retain all eigenvectors with a corresponding eigenvalue at least 1% of the maximum eigenvalue.
- (2) Construct a rectangular matrix, E , consisting of the first-order kernel and the retained eigenvectors weighted by their respective eigenvalues as columns. In addition, the eigenvectors are normalized by the root-mean-square of the input record, although we find that this step has a very minimal effect on the obtained PDMs.
- (3) Perform a singular value decomposition on the rectangular matrix,

$$E=USV^*. \quad (18)$$

The columns of the U matrix form a set of orthonormal basis vectors, and the diagonal of S gives the associated singular values. We retain the five most significant columns of U as the principal dynamic modes.

This procedure yields five forward and five feedback PDMs, for a total of 10, as the second-order kernels always have five significant eigenvectors. This is unsurprising, as five Laguerre basis functions are used in the NARV estimation step to generate the second-order kernel. Moreover, model predictions for the PDM model with five forward and five feedback PDMs are identical to the NARV model predictions, and so the full PDM model retains all the “information” in the NARV model.

Once the PDMs are estimated, they are used as a basis for Volterra kernel expansion, and the general model framework is the same as the NARV. The exogenous and autoregressive inputs are filtered through PDM filterbanks and their outputs are fed to the polynomial expansion, whose coefficients are estimated using the same procedure as for the NARV model (Sec. 2.2.2).

In the interests of model parsimony and interpretability, we are generally interested in models that consider only a subset of all the PDMs and/or a subset of all expansion coefficients, and these are the primary focus of the paper.

2.3. Model reduction: PDM subset selection

We reduce model complexity in a two-step process: (1) identify a subset of PDMs that have the strongest influence on the output, (2) using such a reduced subset of PDMs as a basis, prune the expansion coefficients. Here, we discuss the first. In most previous works, the PDMs have been ranked by eigenvalue or singular value, and some subset based on this ranking is selected (e.g. top four, retain those PDMs with eigenvalues at least 1% of the maximum, etc.). However, in the present work the four least significant singular values are four to five orders of magnitude smaller than the first singular value, and we have found that for any PDM but the first, the singular value is not a reliable indicator of its importance. We suspect that this is related to NARV model estimation errors.

2.3.1. PDM selection—We perform a leave-one-out (LOO) process by which 10 reduced PDM models are estimated, each omitting a single (forward or feedback) PDM. The performance of each reduced model is compared to the full model, and if the performance is found to be significantly inferior as measured by the NRMSE, using the bootstrap procedure detailed in Sec. 2.4.2, the omitted PDM is deemed significant.

The LOO procedure yields a subset of significant PDMs that form the bases for the “LOO-reduced PDM model”. We then use the LOO-reduced model as the baseline model and perform an add-one-in (AOI) procedure, whereby each omitted PDM is added in turn, and the performance of the resulting model is compared to that of the LOO-reduced model. If adding any PDM significantly improves the NRMSE, it is added back to the basis set. The basic rationale for this procedure is that while the LOO step tells us what happens when leaving any one PDM out, it says nothing about leaving two, three, etc. PDMs out, and the AOI step is an attempt to compensate for this. This LOO/AOI procedure may be performed an arbitrary number of times to zero-in on an optimal basis set of PDMs (somewhat analogous to a binary search). We refer to the final model arrived at via this procedure as the “basis-reduced PDM model”.

2.3.2. Model coefficient pruning—In this section, we describe our method for reducing PDM model order by pruning the insignificant expansion coefficients. Then, using our decomposition of the polynomial into various self- and cross-terms (see Fig. 1), retaining only the significant terms allows visualization of the model's most important structural characteristics (see, e.g. Fig. 3 in the Results).

Instead of building the list of significant coefficients from the “bottom-up” by adding successive coefficients and determining if they significantly improve prediction accuracy, as has been done in previous works,²⁶ we prune the coefficients list from the “top-down,” using a LOO approach. That is, we iterate through every coefficient and remove it from the full list, and determine if the “reduced-by-one” model's performance on the testing data-set is significantly diminished. The coefficients of each reduced-by-one model are re-estimated

from the training data-record. This approach minimizes the risk of missing important interactions among terms that may be missed in the bottom-up method.

Under decomposition of the polynomial expansion, the self-forward and self-feedback terms are quadratic and encompass two coefficients. Therefore, we also assess model performance with these coefficients lumped to determine if the two-coefficient self-terms are significant as a whole even if the two coefficients assessed individually do not attain significance.

Finally, given a list of significant coefficients, we re-estimate the coefficients of the “pruned” model and assess model performance. As for PDM basis selection, significance is assessed by the bootstrap procedure detailed in Sec. 2.4.2, using out-of-sample data.

Since coefficient estimation is a multiple linear regression problem, it is also possible to use an F-test to determine if R^2 is significantly reduced by the omission of any coefficient, but we have found that the F-test tends to retain an excessive number of coefficients; the bootstrap method is more flexible and useful.

2.4. Performance evaluation

2.4.1. Metrics—We use the NRMSE of the continuous membrane potential as the primary metric for model evaluation, PDM basis set reduction, and coefficient pruning. We are also interested in the comparative ability of models to predict APs in a binary (all-or-nothing) sense. Therefore, we convert the continuous membrane potential to a spike-train by imposing a threshold for detection θ_D . Anytime membrane potential goes above θ_D from below, an AP is recorded. We also impose a 4 ms refractory period, so that no additional AP may be recorded within 4 ms of an AP detection.

To compare spike trains, we use two statistics: (1) the coincidence factor without replacement, $\Gamma^{15,28}$ and (2) the K -statistic for the two-sample Kolmogorov–Smirnov (KS) test comparing the interspike time distributions. The coincidence factor is given as

$$\Gamma = \left(\frac{N_{coinc} - \langle N_{coinc} \rangle}{0.5 (N_{data} + N_{model})} \right) \left(\frac{1}{\Lambda} \right), \quad (19)$$

where

$$\langle N_{coinc} \rangle = \frac{N_{data} N_{model}}{K}, \quad (20)$$

$$\Lambda = 1 - \frac{N_{model}}{K}. \quad (21)$$

The total AP counts for the data and model output are N_{data} and N_{model} . An AP in the model output and actual data are considered coincidental if within $\pm \Delta$; we set $\Delta = 3$ ms. Counting all such coincidences, with the restriction that no predicted AP may correspond with more than one actual AP and vice versa, yields N_{coinc} . The parameter $\langle N_{coinc} \rangle$ gives the expected number of coincidences generated by a Poisson process with the same AP frequency as the model, with the data record divided into K bins of length 2Δ . Subtracting $\langle N_{coinc} \rangle$ from

N_{coinc} ensures that Γ will tend toward 0 when all coincidences occur purely by chance. Dividing by Λ normalizes the measure such that $\Gamma = 1$ when the predicted and actual spike-trains are in perfect agreement, i.e. $N_{\text{coinc}} = N_{\text{data}} = N_{\text{model}}$.

The K -statistic gives the maximum difference between the empirical cumulative distribution functions of the interspike times.

For the simulated data, we fix $\theta_D = 50$ mV. Choosing θ_D for model predictions is somewhat problematic, as variations in predicted AP morphology may result in clearly recognizable APs that fail to meet a given θ_D . Highly pruned models, in particular, tend to give lower AP peaks. To make our model comparisons as fair as possible, we optimize θ_D for each individual model with respect to the statistic of interest (Γ or K -statistic). This allows us to compare the best possible model predictions against each other.

We also consider as secondary metrics the true positive rate (TPR) and false positive rate (FPR), which are calculated following binning of spikes into 2 ms wide bins as:

$$\begin{aligned} TPR &= \frac{\text{Number of true positives}}{\text{Number of actual APs}}, \\ FPR &= \frac{\text{Number of false positives}}{\text{Number of bins lacking an AP}}. \end{aligned}$$

A predicted AP is considered a true positive if it occurs within a 1 bin margin of an actual AP. In addition, no predicted AP is allowed to coincide with more than one actual AP, and no actual AP may coincide with more than one predicted AP, i.e. no double-counting is allowed. Any AP which is not deemed a true positive is counted as a false positive. Optimizing Γ tends to result in a TPR and FPR that are slightly lower than when optimizing the K -statistic. That is, Γ has a slight “preference” for decreasing the FPR at the expense of the TPR relative to the K -statistic.

We also generate receiver operating characteristic (ROC) curves for θ_D , which plot TPR against FPR and each point represents a particular value of θ_D . The ROC curve may also be used to select the optimal θ_D , with the point closest to the upper-left corner as the most typical choice.²⁶ However, because FPR is extremely low for all feasible values of θ_D (as almost all bins lack an AP), we do not use the ROC curves to choose θ_D , but only to visualize comparative model performance.

Using the NRMSE as the test statistic for PDM basis set reduction and coefficient pruning guards against any sensitivity of spike-train similarity metrics on θ_D . Finally, note that θ_D is *not* the same parameter as θ , which determines when the model autoregressive input is nonzero and is always fixed at 4.5 mV. While using the NRMSE guards against errors or biases resulting from our automated spike-train extraction, it has the disadvantage of being highly sensitive to small temporal offsets between spikes, and other metrics for comparing voltage tracings that do not rely on spike extraction have been employed. For example, a popular method developed by LeMasson and Maex²⁷ considers the overlap between model and data trajectories in the dV/dt versus V phase-plane. Nevertheless, the NRMSE has proven an adequate test statistic, and we restrict ourselves to this metric in this work.

2.4.2. Bootstrap procedure for estimating variance—To compare model performances, we must estimate the variance of our performance metrics. We calculate the bootstrap distribution of the mean of a statistic on the testing data as follows. A given model is run for the 10 8192 ms testing data-sets, and the model results are concatenated into the full prediction record. This record is divided into 80 snippets of 1024 ms each. The metric of interest (e.g. NRMSE) is then calculated for each snippet, giving 80 estimates. Then, 80 samples are drawn with replacement from this set of estimates and averaged to give an estimate of the mean. 10,000 such re-samplings are performed to yield the bootstrap distribution of the mean.

To, for example, determine if the NRMSE for a full and reduced model varies at the 0.05 significance level, we calculate the bootstrap distribution of the NRMSE difference. If the 95% confidence interval does not include zero, then the reduced model is considered to have a significantly different NRMSE. We assess significance at the 0.01, 0.05, 0.10, and 0.20 significance levels using the two-sided confidence interval. This procedure is performed on out-of-sample (testing) data, and so determines the variance of the *predictive* performance.

3. Results

3.1. Estimated principal dynamic modes—The five estimated forward and feedback PDMs for the baseline H–H model are given in Fig. 2. In the following section, we find two forward and four feedback PDMs to be significant, and re-label them, in order of significance, as X1, X2, Y1, Y2, Y3, and Y4. To get insight into their functional role, we treat these as linear filters and examine their response to pulse and step inputs. The major forward PDM, X1, is a simple integrator, while X2 has the characteristics of both a differentiator and a slow or delayed integrator.

We also examine the filter outputs of the Y1 and Y2 in response to an input pulse. Since the autoregressive component is typically only nonzero for brief periods corresponding to APs, a pulse is a much more appropriate interrogator than a step input. Based on the LOO results in Sec. 3.1.1, Y1 and Y2 are the two most influential feedback PDMs; Y1 appears to be principally responsible for the AP waveform, while Y2 has a delayed integrator characteristic and gives the slower afterpotential. The Y3 and Y4 PDMs also have a slightly delayed response to the pulse and appear to contribute to the afterpotential or refractory period, but have rather complex responses.

3.1.1. PDM basis reduction—We perform the alternating leave-one-out/add-one-in (LOO/AOI) procedure described in Sec. 2.3.1, using the NRMSE as the test statistic. Using a significance level of 0.10 we get rapid convergence to the X1–2, Y1–4 PDM model, which we refer to as the 2-4 basis set. This model has 27 free expansion coefficients, contra the 65 of the full PDM model, and we refer to it as the 27-P model hereafter.

3.2. Pruned PDM models

We apply the LOO coefficient pruning algorithm to several PDM models. The difference between the full and each reduced-by-one model, $\text{NRMSE}_{\text{diff}}$, is used as the test statistic, and two methods for deciding to prune each coefficient are proposed:

- (1) If $\text{NRMSE}_{\text{diff}}$ is significantly different from 0, then the coefficient is retained. Otherwise, it is pruned. Results are very insensitive to the prescribed significance level, being identical for significance levels of 0.01, 0.05, and 0.10, and usually identical for the 0.20 significance level.
- (2) An absolute threshold is set, and if $\text{NRMSE}_{\text{diff}}$ is greater than this threshold, the coefficient is retained.

The first method yields pruned models with fewer coefficients and performance that is good but slightly inferior to the full PDM model. The second method, for a judiciously chosen threshold, yields pruned models with several more coefficients but performance comparable to the full model, as explored further in Sec. 3.2.3. Unless otherwise specified, all pruned PDM models are determined using Method 1. We also rank the retained coefficients by $\text{NRMSE}_{\text{diff}}$ to give an indication of the relative importance of each term.

3.2.1. Pruned 2-4-basis PDM model: the 8-P model—Applying Method 1 with a significance level of 0.10 to the 27-P PDM model, we obtain the eight-parameter (8-P) model, depicted in Fig. 3. There is a single self-forward term in X_1 and a single self-feedback term in Y_1 , both of which are *linear* (first-order). Since X_1 is integrative, we conclude that the forward separable component of the model is linear and integrative. The model is dominated by cross-terms, most of which are cross-forward-feedback terms. The X_1 and Y_1 PDMs are clearly of central importance, together contributing to seven out of eight terms.

3.2.2. Pruned full-basis PDM model: the 4-P model—We apply the pruning algorithm to the full 65-P PDM model. Perhaps surprisingly, the full-pruned PDM models are markedly inferior to the 2-4-pruned PDM models. Pruning of the full model leaves four coefficients and three PDMs, X_1 , Y_1 , and Y_2 , as shown in Fig. 4. This model has linear forward terms in X_1 and Y_1 , and X_1 – Y_2 and Y_1 – Y_2 cross-terms. These terms are the same as the top four terms of the 8-P pruned model, as ranked by $\text{NRMSE}_{\text{diff}}$.

From this, we conclude that direct application of pruning to the full PDM/NARV model reliably identifies the most essential model components. However, to estimate those components of significant but not overwhelming importance, it is important to reduce the PDM basis sets first. The superior predictive performance obtained when reducing the PDM sets before pruning justifies this conclusion.

3.2.3. N-coefficient models—To determine how model performance changes with the number of expansion coefficients included, for a given PDM basis set (i.e. either the full or the 2-4 basis sets), we construct a set of “ N -coefficient” models, where each coefficient of the model is ranked by its $\text{NRMSE}_{\text{diff}}$ as determined by the LOO procedure described above, and the first N coefficients are included. We do this for the full basis set (65-P model) and the 2-4 basis set (27-P model); Fig. 5 plots the bootstrap mean and 95% confidence intervals for the NRMSE , Γ , TPR, and FPR as functions of N .

As can be seen from the figure, NRMSE (and Γ) actually improves as the first few coefficients are omitted, i.e. the $N = 20$ – 26 models perform better than the full $N = 27$ model

under the 2-4 basis set, and under the full basis set, the $N = 50$ –64 models are all significantly better than the $N = 65$ model. This is likely related to overfitting. As N is further reduced, model performance degrades gradually and slightly until there is a dramatic and abrupt drop in performance at either seven (2-4 basis set) or six (full basis set) coefficients.

We conclude that a small number of model coefficients are actually detrimental to performance, the majority have a minor or nonexistence effect on performance, and a small number, between 6 and 10, are responsible for almost all predictive ability. Moreover, the NRMSE performance curve for the 2-4 basis set is always significantly below that of the full basis set (with the sole exception of $N = 18$), confirming the value of reducing the basis set before coefficient pruning.

The location of the 8-P pruned model identified by the LOO pruning procedure on the N -coefficient curve shows that the procedure can identify the (nearly) minimal model that retains good performance. Alternatively, one could simply select the desired model from the N -coefficient model curve, which explicitly displays the trade-off between parsimony and performance.

3.2.4. Predicted time-series membrane potential and spike trains—Example 500 ms snippets of the membrane potential tracings for the 27-P, 8-P, and 4-P models are given in Fig. 6. On visual inspection, the 27-P model predictions are essentially identical to the full 65-P model predictions, which is why we have omitted the latter from the figure. The pruned models (8-P and 4-P) typically capture the AP waveform less precisely and predict APs with a slightly lower peak amplitude. In the case of the 8-P model, this has only a minimal effect on the ability to predict APs in the binary sense. This is not the case with the 4-P model.

3.2.5. Comparative performance: coincidence factor and NRMSE—Figure 7 shows the bootstrap distributions of the mean Γ , NRMSE, TPR, and FPR for the full 65-P, 27-P, 8-P, and 4-P models. These distributions are determined from the full testing data set and represent out-of-sample results. PDM basis set reduction results in a significant increase in the NRMSE ($p < 0.05$) and significant decrease in the TPR ($p < 0.05$), but no significant difference in the coincidence factor or FPR. Pruning leads to a reduction in Γ that achieves statistical significance relative to the full model ($p < 0.05$), but no reduction in TPR and an increase in FPR ($p < 0.01$). Compared directly, the 27-P and 8-P models do not differ significantly in NRMSE or Γ . The 4-P model is clearly inferior in all metrics, and its inclusion in the figure helps demonstrate the very similar performance of the first three models in an absolute sense.

3.2.6. Comparative performance: ROC, Γ , and K-statistic curves—Figure 8 displays ROC curves for the full 65-P, 27-P, 8-P, and 4-P models. There is only a slight degradation in performance, as measured by the area-under-the-curve (AUC), moving from the 65-P to the 8-P model, while the 4-P model is markedly inferior.

We also mark, on the ROC curves, the locations of the optimal thresholds for spike detection, θ_D , with respect to the coincidence factor and K -statistic. We find that both

metrics give similar thresholds, but they favor a higher TPR and FPR compared to the normalized closest point to the upper left.

Figure 9 gives the coincidence factor, Γ , and K -statistic as function of θ_D . Both un-pruned models (65-P and 27-P) are relatively insensitive to θ_D over a fairly broad range, while pruning both lowers the optimal θ_D and narrows the range of good performance. Results for Figs. 8 and 9 are for the full testing data-set.

3.2.7. Interspike time histograms—We generate interspike time histograms for the full 65-P, 27-P, and 8-P models on the full testing data-set. These are compared to the actual interspike time histogram, as shown in Fig. 10. We also determine the empirical cumulative distribution function (CDF) for each such histogram, and we perform a two-sample KS test on each histogram pair, which tests the null hypothesis that the two data-sets are drawn from the same distribution. Under the KS test, the data and the model predictions all are drawn from different distributions, although the data to 65-P model comparison is of borderline significance at $p = 0.0301$. The 65-P model and 27-P interspike time histograms are nearly identical, and the KS test fails to reject the null hypothesis at $p = 0.6945$. The 8-P model histogram shows the poorest agreement with the data and is significantly different from all the other histograms (the 4-P model is much worse and is omitted).

3.2.8. Volterra kernels for full and reduced models—Comparing the Volterra kernels reconstructed from the full and reduced PDMs allows direct visualization of how greatly the models vary. We compare the full 65-P, 27-P, 8-P, and 4-P models. As shown in Fig. 11, the first-order forward and feedback kernels are essentially the same for all four models. The second-order forward self-kernel disappears under both pruned models (8-P and 4-P), while the second-order feedback self-kernels are also all similar, with the only major differences occurring at one or two lags (not shown).

The major difference between the 4-P model and the others is in the second-order cross-kernel. The high-frequency component, which can be accounted for by the $X1-Y1$ cross-term, is missing from the cross-kernel of the 4-P model. Therefore, this cross-term is demonstrated to be highly significant with respect to kernel reconstruction.

The good agreement in kernel morphology between the full model, with 65 expansion coefficients, and the 8-P model confirms that the eight retained terms capture most of the system structure.

3.2.9. Linear minimal model—Pruning reveals the PDM model structure to be dominated by an integrative forward PDM with a linear ANF and a feedback PDM that represents the early AP waveform, also with a linear ANF (Figs. 3 or 4). Presumably, the nonlinear cross-terms modulate AP shape and are important contributors to the afterpotential and refractoriness. We have tested these assumptions by examining the performance of a model retaining only the first forward and feedback PDMs and the associated linear ANFs, which we refer to as the linear minimal model.

On the full testing dataset, the bootstrap means for the coincidence factor, Γ , and the NRMSE, are 0.41 and 0.14, respectively, which are markedly inferior to both the 8-P and 4-P models, demonstrating the importance of cross-terms to overall performance. Sample membrane potential tracing for the 8-P and linear minimal models are given in Fig. 12. Figure 12 also shows the response of these two models to a 1 ms current pulse of 20 μA , showing that the cross-terms are also essential to capturing the AP waveform.

We have also replicated the numerical experiment reported in Ref. 2, where pairs of 1 ms current pulses are given in sequence, with a progressively wider interval between the two, to determine if the reduced PDM models correctly predict the existence of a relative and absolute refractory period. While both the 8-P and 4-P PDM models perform reasonably well (although they do underestimate the duration of the refractory periods), the linear minimal model fails completely to demonstrate refractoriness.

3.3. Altered channel dynamics

Changes in the underlying system's ion channel dynamics are necessarily reflected in data-derived Volterra-style models. We have examined how changing the maximum sodium and potassium conductances, $\bar{g}_{Na}$ and $\bar{g}_K$, respectively, affects the estimated PDM morphology and reduced/pruned model structure. A number of medically useful drugs, including local anaesthetics and many antiarrhythmics, and many toxins, most notably tetrodotoxin (TTX), inhibit sodium conductance (per Paracelsus, “all substances are poisons...The right dose differentiates a poison and a remedy”).

TTX inhibition of sodium channels in giant squid axon is well modeled by decreasing $\bar{g}_{Na}$ in the H–H equations,^{29,30} and does not appear to affect the temporal dynamics of the sodium channel, represented by the m and h variables in the H–H equations.³⁰ Therefore, we have generated training and testing datasets with $\bar{g}_{Na}$ and $\bar{g}_K$ varying from 60 to 240 $\mu\text{A cm}^{-2}$ and from 12 to 84 $\mu\text{A cm}^{-2}$, respectively.

We have found that, in general, the PDMs of the modified systems resemble those of the original but the ranks of the singular values of the most similar PDMs are not necessarily equal. To facilitate comparison, for each conductance-modified forward (feedback) PDM we find the correlation coefficient between it and the five original PDMs, and the most similar PDMs are paired. In general, the paired PDMs are quite similar. Given this, we choose to impose a common basis set across models so that they can be compared more easily within the coefficient space, and we use the PDM basis set estimated for the original, unaltered H–H system as the “universal PDM basis”. For systems with altered ion conductance, use of the universal basis in lieu of a system-specific PDM basis only slightly decreases performance.

Under coefficient pruning, the resulting coefficient structure tends to be similar to that obtained for the original H–H system, and in particular, the four coefficients of the original P-4 model tend to be identified as the most essential. We have tried examining the pruned model structures for different conductance-modified model, and while they clearly vary, it is difficult to gain insight when considering seven or eight parameters. Therefore, we impose

the basic 4-P model structure, consisting of X1 and Y1 forward terms, and Y1–Y2 and X1–Y2 cross-terms (see Fig. 4), and determine how (and if) the values of the four coefficients change systematically with $\bar{g}_{Na}$ and $\bar{g}_K$. As shown in Figs. 13 and 14, each term exhibits a clear trend.

The most significant trends are in the cross-terms: as sodium conductance increases these terms diminish, while they are enhanced by increases in potassium conductance. These terms contribute principally to the afterpotential and refractoriness, and it is quite sensible that increased potassium conductance should enhance refractoriness and sodium conductance diminish it. Indeed, the effects of increasing sodium and potassium conductance on the four coefficients of the 4-P model are nearly inverse images of each other, but this antisymmetry is broken in that increasing either conductance decreases the magnitude of X1.

3.4. Model reduction applied to the NARV model with Laguerre basis

The NARV model, in which the Volterra kernels are expanded on a Laguerre basis, is mathematically equivalent to the PDM model structure, and the proposed basis set reduction and coefficient pruning methods may be directly applied to the NARV model. We have found that, in contrast to the PDM model, much less parsimony can be achieved through direct reduction of the NARV model, and the pruned NARV models are not interpretable.

Basis set reduction of the NARV model results in a 44-coefficient model that has performance comparable to the 27-P PDM model that results from the un-pruned 2-4 PDM basis. Pruning of the basis-reduced NARV model results in extremely poor performance (not shown), while pruning of the full NARV model gives the 10-coefficient model (with eight associated DLFs) depicted in Fig. 15. This model is difficult to interpret, does not suggest that the membrane acts principally as a leaky integrator of current in the subthreshold regime, and the AP waveform and afterpotential are not represented as in the pruned PDM-based model. This model, which consists of eight basis functions and 10 coefficients, has performance that is significantly inferior to the more compact 8-P model and only comparable to the much more interpretable 4-P PDM model (which consists of only three PDM basis functions and four coefficients). Thus, we demonstrate the importance of employing the PDM basis set with respect to both model reduction and interpretation.

4. Discussion

This study has explored the use of PDM analysis of the system described by the H–H equations to achieve parsimonious PDM-based models of predictive capability comparable to the full Nonlinear Autoregressive Volterra model representation that was recently published.² We find that a two-step model reduction procedure is advisable for this purpose, whereby a set of PDMs is first identified that efficiently represents the equivalent Volterra model, and subsequent pruning of the PDM-based model terms yields models of dramatically reduced complexity with comparable predictive capability. This procedure identifies an eight-parameter (“8-P”) model, representing an eightfold reduction relative to the full NARV model, and a threefold reduction relative to the differential equations representation of the H–H system.

We have demonstrated that the proposed methodology greatly reduces PDM-based model complexity and may facilitate model interpretation. Reduction to the 2-4 PDM basis gives a model (27-P model) consisting of six PDMs and 27 free parameters, compared with 10 PDMs and 65 free parameters for the full model. Its performance is only slightly inferior to the full model with respect to prediction NRMSE, and comparable with respect to most measures of spike-train prediction, such as the coincidence factor, Γ , the ROC curves, the interspike-time histograms, and the curves describing Γ and K -statistic dependence on the threshold for spike detection (see Fig. 9).

Further pruning of the 27-P model yields a compact model with only eight free parameters (8-P model) and whose performance is only slightly inferior to that of the full model. Moreover, with only eight free parameters most kernel structure is preserved upon reconstruction of the Volterra kernels (see Fig. 11). We conclude that PDM-based modeling with basis reduction is advisable in this system, as it reduces model complexity by one-half to two-thirds at essentially no performance cost. Further pruning also appears advisable, as it leads to greater model parsimony with only a slight degradation of performance.

Examination of the form of the obtained forward PDMs for the H–H equations/system reveals that X1 is a “leaky integrator” over an input past-epoch of approximately 5 ms, and X2 is a “finite-bandwidth differentiator” (see Fig. 2). Marmarelis³¹ previously proposed a model for single-neuron operation comprising two “neuronal modes” that are analogous to the X1 and X2 leaky-integrator and slow-differentiator PDMs, followed by a static nonlinearity and a threshold-trigger operator. The present analysis seems to corroborate that postulate, but only with regard to the forward branch of the PDM-based model of the H–H equations (suitable for the sub-threshold operation of this system). However, in the supra-threshold operation of this system, our analysis reveals a feedback PDM (Y1) responsible for the waveform of the generated AP, and three more feedback PDMs that partake in the supra-threshold dynamics of the system via modulatory influences principally upon the output of the first forward and first feedback PDMs (see Fig. 3), expressed in the PDM-based model as additive pair-product terms.

Further pruning reveals that the most significant of these modulatory influences are the ones exerted by the second feedback PDM (see Fig. 4) which exhibits integrative characteristics over roughly 15 ms. Moreover, under pruning the X2 PDM is preserved only in a forward-feedback term, indicating that the rate of current injection also influences post-AP dynamics. We note that the fourth feedback PDM exhibits dynamic characteristics of first-order differentiation. Future studies will seek to explore the relation of these PDMs with the specific ion-channel mechanisms of this system.

The most highly pruned model, shown in Fig. 4, captures most of the essential system characteristics, and we give the following interpretation. The dominant forward PDM of the H–H system is X1, which is integrative, with a linear ANF. In the sub-threshold regime, it is the only active model component, and thus the sub-threshold model acts purely as a leaky integrator of recent current input, with a memory of roughly 5 ms, consistent with the widely used integrate-and-fire model. However, X1 is not a simple exponential waveform, but appears to have three different time constants. The fact that the NARV model retains two

separable forward bases (Fig. 15) under model reduction, one of which is an exponential, suggests that this is not simply an artifact of model estimation. In the supra-threshold regime, the model has more complex structure including a linear feedback term and two pair-product terms between the second feedback PDM (Y2) and the first forward (X1) and first feedback (Y1) PDM. The Y1 PDM is responsible for the early features of the AP waveform, and also contributes to the afterpotential. The pair-product terms (cross-terms) of the model contribute to peri-AP phenomena and have an effect on the refractory period. Reconstruction of the Volterra kernels (Fig. 11) also suggests that the X1–Y1 term is an important contributor to the high frequency characteristics of the Volterra cross-kernel.

It must be noted that, to generate synthetic data, we have used Hodgkin and Huxley's original parameter set,¹ which is calibrated for giant squid axon at 6.3°C. Action potentials generated at body temperature are expected to be narrower, and hence training the NARV and PDM models under such a condition would likely necessitate a smaller sampling interval. Furthermore, we may expect variation in the basic PDM model structure; at the very least a reduction in the time constant for the first feedback PDM, Y1, would be expected.

We have also performed a preliminary analysis of how changes in underlying channel dynamics affect the coefficients of the maximally reduced 4-P model. We have observed that changes in the maximum sodium and potassium conductance affect the form of the PDMs only mildly, which maintain their main functional characteristics (e.g. integrative, differentiating, etc.). This, and the difficulty of comparing waveforms versus comparing scalars, motivates the strategy of fixing a universal PDM basis and coefficient set, and examining system variability within coefficient space. Using the 4-P model as the universal framework, changes in the sodium and potassium conductance consistently affect the X1, Y1, Y1–Y2, and X1–Y2 terms of the 4-P model (see Figs. 13 and 14). These changes are most marked in the Y1–Y2 and X1–Y2 cross-terms, both of which increase in magnitude with potassium conductance, but decrease in magnitude with sodium conductance.

Application of the two-step model reduction procedure directly to the NARV model demonstrates the importance of using PDMs as a basis. Model reduction under the PDM basis leads to more compact models with better predictive capability. While performing a basis-reduction step before pruning almost uniformly improves model performance when using a PDM basis (see Fig. 5), this does not appear to be the case when using the Laguerre basis of the NARV model, and in fact is detrimental to performance (not shown). Moreover, the pruned full NARV model (shown in Fig. 15) has no obvious interpretation in terms of the functional characteristics of the H–H system, in stark contrast to the pruned PDM models.

In the current work, we have largely restricted our exploration of PDM model dynamics to white-noise inputs and accurate prediction of AP generation by the H–H membrane. The H–H model is well known to exhibit rich dynamics, the most famous and basic being threshold phenomena, refractoriness, and limit cycle behavior. The NARV model was previously shown to produce these behaviors,² and we have found that the 27-P, 8-P, and 4-P PDM models all yield these basic dynamics as well (results not shown). The H–H model also

exhibits more complex dynamics³² such as hysteresis, bistability, and the phenomenon of anode-break excitation. However, it is beyond the scope of the current work to study such behavior under the PDM framework.

As demonstrated in Fig. 9, the optimal threshold for spike detection is not altered by PDM basis reduction, but is markedly affected by coefficient pruning. It is, therefore, important to optimize metrics of spike-train similarity, such as Γ , for each individual model to get a fair assessment of its performance. We have also found that, for pruned models, the optimal threshold for detection, θ_D , changes with input power; namely, lower input power leads to lower detection thresholds for pruned PDM-based models (results not shown). While not emphasized in this work, we have previously found the NARV model trained with high-power input is applicable to lower power out-of-sample inputs,² and we have similarly found a single high-power input adequate to train PDM-based models applicable to lower power inputs (results not shown).

In conclusion, our results on model reduction indicate that many terms of the full PDM-based model of the H–H equations make insignificant contributions to model predictive performance, as assessed by all prediction metrics used (see Figs. 7–11) and the reconstructed equivalent Volterra kernel estimates (see Fig. 11). Moreover, Fig. 5 suggests overfitting when all coefficients are included, and only 6–10 coefficients are necessary to capture the main H–H dynamics. This finding suggests that PDM-based modeling with appropriate pruning can achieve parsimonious Volterra-equivalent models of the H–H equations with excellent predictive capability.

Acknowledgments

This work was supported in part by the Biomedical Simulations Resource at the University of Southern California under NIH grant P41-EB001978.

References

1. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Phys.* 1952; 117(4):500–544.
2. Eikenberry SE, Marmarelis VZ. A nonlinear autoregressive Volterra model of the Hodgkin–Huxley equations. *J. Comput. Neurosci.* 2013; 34(1):163–183. [PubMed: 22878687]
3. Volterra, V. *Theory of Functionals and of Integro-Differential Equations*. Dover Publications; New York: 1930.
4. Volterra V. Sopra le funzioni che dipendono da altre funzioni. *R. C. Accad. Lincei.* 1887; 3:97–105, 141–146, 153–158.
5. Volterra V. Sopra le funzioni permutabili. *R. C. Accad. Lincei.* 1910; 5(19):425–437.
6. Marmarelis, VZ. *Nonlinear Dynamic Modeling of Physiological Systems*. Wiley-IEEE Press; Hoboken: 2004.
7. Marmarelis VZ. Identification of nonlinear biological systems using Laguerre expansions of kernels. *Ann. Biomed. Eng.* 1993; 21(6):573–589. [PubMed: 8116911]
8. Guttman R, Feldman L, Lecar H. Squid axon membrane response to white noise stimulation. *Biophys. J.* 1974; 14(12):941–955. [PubMed: 4429772]
9. Guttman R, Feldman L. White noise measurement of squid axon membrane impedance. *Biochem. Biophys. Res. Commun.* 1975; 67(1):427–432. [PubMed: 1201032]
10. Bryant HL, Segundo JP. Spike initiation by transmembrane current: A white-noise analysis. *J. Physiol.* 1976; 260(2):279–314. [PubMed: 978519]

11. Guttman R, Grisell R, Feldman L. Strength-frequency relationship for white noise stimulation of squid axons. *Math. Biosci.* 1977; 33(3):335–343.
12. Buño W, Bustamante J, Fuentes J. White noise analysis of pace-maker-response interactions and non-linearities in slowly adapting crayfish stretch receptor. *J. Physiol.* 1984; 350(1):55–80. [PubMed: 6747858]
13. Korenberg MJ, French AS, Voo SKL. White-noise analysis of nonlinear behavior in an insect sensory neuron: Kernel and cascade approaches. *Biol. Cybern.* 1988; 58(5):313–320. [PubMed: 3382702]
14. Bustamante J, Buño W. Signal transduction and nonlinearities revealed by white noise inputs in the fast adapting crayfish stretch receptor. *Exp. Brain Res.* 1992; 88(2):303–312. [PubMed: 1577104]
15. Kistler WM, Gerstner W, van Hemmen JL. Reduction of Hodgkin–Huxley equations to a single-variable threshold model. *Neural Comput.* 1997; 9(5):1015–1045.
16. Lewis ER, Henry KR, Yamada WM. Essential roles of noise in neural coding and in studies of neural coding. *BioSystems.* 2000; 58(1):109–115. [PubMed: 11164637]
17. Takahata T, Tanabe S, Pakdaman K. White-noise stimulation of the HodgkinHuxley model. *Biol. Cybern.* 2002; 86(5):403–417. [PubMed: 11984654]
18. Jolivet R, Rauch A, Lscher HR, Gerstner W. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *J. Comput. Neurosci.* 2006; 21(1):35–49. [PubMed: 16633938]
19. Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W. A benchmark test for a quantitative assessment of simple neuron models. *J. Neurosci. Methods.* 2008; 169(2):417–424. [PubMed: 18160135]
20. Watanabe A, Stark L. Kernel method for nonlinear analysis: Identification of a biological control system. *Math. Biosci.* 1975; 27(1):99–108.
21. Ogura, H. Estimation of Wiener kernels of a nonlinear system and a fast algorithm using digital Laguerre filters. 15th NIBB Conf.; Okazaki, Japan. 1985. p. 14-62.
22. Marmarelis VZ, Shin DC, Song D, Hampson RE, Deadwyler SA, Berger TW. Nonlinear modeling of dynamic interactions within neuronal ensembles using Principal Dynamic Modes. *J. Comput. Neurosci.* 2013; 34(1):73–87. [PubMed: 23011343]
23. Knight BW. Dynamics of encoding in a population of neurons. *J. Gen. Physiol.* 1972; 59(6):734–766. [PubMed: 5025748]
24. Jolivet R, Lewis TJ, Gerstner W. Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *J. Neurophysiol.* 2004; 92(2):959–976. [PubMed: 15277599]
25. Marmarelis VZ. Modeling methodology for nonlinear physiological systems. *Ann. Biomed. Eng.* 1997; 25(2):239–251. [PubMed: 9084829]
26. Zanos TP, Courellis SH, Berger TW, Hampson RE, Deadwyler SA, Marmarelis VZ. Nonlinear modeling of causal interrelationships in neuronal ensembles. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2008; 16(4):336–352. [PubMed: 18701382]
27. LeMasson, G.; Maex, R. Introduction to equation solving and parameter fitting. In: De Schutter, E., editor. *Computational Neuroscience: Realistic Modeling for Experimental-ists*. CRC Press; London: p. 1-24.
28. Gerstner W, Naud R. How good are neuron models? *Science.* 2009; 326(5951):379–380. [PubMed: 19833951]
29. Narahashi T, Moore JW, Scott WR. Tetrodotoxin blockage of sodium conductance increase in lobster giant axons. *J. Gen. Physiol.* 1964; 47(5):965–974. [PubMed: 14155438]
30. Takata M, Moore JW, Kao CY, Fuhrman FA. Blockage of sodium conductance increase in lobster giant axon by tarichatoxin (tetrodotoxin). *J. Gen. Physiol.* 1966; 49(5):977–988. [PubMed: 5961361]
31. Marmarelis VZ. Signal transformation and coding in neural systems. *IEEE Trans. Biomed. Eng.* 1989; 36(1):15–24. [PubMed: 2646209]
32. Guttman R, Lewis S, Rinzel J. Control of repetitive firing in squid axon membrane as a model for a neuroneoscillator. *J. Physiol.* 1980; 305(1):377–395. [PubMed: 7441560]

Fig. 1.

Example PDM model architecture for two forward and feedback PDMs. Each PDM-defined filter-bank receives its respective input and generates PDM outputs that are transformed by polynomial static nonlinearities and summed to form the model output. The static nonlinearity is decomposed into (a) self-forward, (b) cross-forward, (c) self-feedback, (d) cross-feedback, and (e) cross-forward-feedback terms. For the second-order architecture, the self-terms are quadratic, while the cross-terms are bilinear (i.e. a scalar multiple).

Fig. 2.

The five forward (top row) and feedback (bottom row) PDMs ordered by significance. The two forward (X1, X2) and four feedback (Y1, Y2, Y3, Y4) PDMs found to be significant are bolded.

Fig. 3. Pruned 8-P PDM model structure obtained under coefficient pruning of the 27-P PDM model (i.e. the 2-4 basis set), using Method 1 with a significance level of 0.10. The coefficients are ranked by $\text{NRMSE}_{\text{diff}}$.

Fig. 4.

The four surviving terms under direct pruning of the full 65-P PDM model, yielding the maximally reduced 4-P model. Note that these coefficients are the same four identified as most essential in the 8-P model.

Fig. 5.

Performance metrics for N -coefficient PDM models under the full and 2-4 PDM basis sets. The bold lines give the bootstrap mean, and the shaded regions enclosed by dotted lines indicate the 95% CIs. The asterisks mark the location of the 8-P model. The coefficients of the 8-P model are selected by significance testing on all coefficients of the 2-4 basis set, and are found to be the same as those retained by $\text{NRMSE}_{\text{diff}}$ ranking. Performance is assessed using the full out-of-sample testing data-set.

Fig. 6.

Out-of-sample membrane potential predictions for, from top to bottom, the 27-P, 8-P, and 4-P models. As predictions for the full 65-P and 27-P models are nearly identical in appearance, we omit results for the full model.

Fig. 7.

The bootstrap distributions of the mean Γ , NRMSE, TPR, and FPR for the full 65-P, 27-P, 8-P, and 4-P models run on the full testing data-set. An asterisk indicates the mean is significantly different from that of the full model at the 0.05 significance level, while an “x” indicates significance at the 0.01 level.

Fig. 8.

(Color online) The four panels on the left show the individual ROC curves for the four models. The marker meanings follow: (1) the red square indicates the point closest to the upper left normalized by the maximum TPR and FPR, (2) the green diamond shows the location of the optimal θ_D with respect to Γ , and (3) the black triangle indicates the optimal θ_D with respect to the K -statistic. The right panel plots all ROC curves together.

Fig. 9.

The left panel gives the coincidence factor, Γ , as a function of the threshold for spike detection, θ_D , and the right panel shows how the K -statistic varies with θ_D . The ranges of good performance are similar between the two metric, and the optimal θ_D values are also quite close. This figure demonstrates that the predicted spike-trains under the nonpruned models are insensitive to θ_D over a broad range, while the pruned models have much narrower θ_D tuning curves. This makes sense, as the principal effect of coefficient pruning is to degrade the model's ability to recapitulate the fine details of the AP waveform.

Fig. 10.

The interspike time histograms for the full testing data-set and those predicted by the full 65-P, 27-P, and 8-P models are given on the left. The empirical CDFs derived from these histograms are plotted on the right. The CDFs for the former three histograms are nearly indistinguishable.

Fig. 11.

Selected Volterra kernels reconstructed from full and reduced PDM models. The left part of the figure gives the first-order forward and feedback kernels, and the right portion gives the second-order cross kernels (with the x - and y -axes having units of delay in ms).

Fig. 12.

Results from the linear minimal model. The left side of the figure gives a 400 ms example of out-of-sample membrane potential predictions for the 8-P (top) versus the linear minimal model (bottom). The right side displays the responses of these two models to a 1 ms pulse of 20 μ A of current. Thus it is demonstrated that the cross-terms excluded from the linear minimal model are essential to accurately representing the AP waveform.

Fig. 13.

Values of the estimated X1, Y1, Y1-Y2, and X1-Y2 coefficients under the 4-P model framework as a function of the underlying H-H system's maximum sodium conductance, $\bar{g}_{\text{Na}}$. Increasing $\bar{g}_{\text{Na}}$ suppresses Y1-Y2 and X1-Y2, diminishing the refractory period and easing AP firing. While the trends in X1 and Y1 are clear, they are relatively small in absolute value. The location of the standard H-H $\bar{g}_{\text{Na}}$ is marked with an asterisk.

Fig. 14.

The counter-trends in X1 and Y1 appear to have the net effect of enhancing the “late” activity of Y1 in response to current input as $\bar{g}_K$ increases, while having little effect on the early AP waveform. Increasing in $\bar{g}_K$ greatly increases the magnitude of the Y1-Y2 and X1-Y2 terms to suppress AP firing; this suppression is active immediately following AP firing, and enhances refractoriness to further firing in response to a current stimulus. The location of the standard H-H $\bar{g}_K$ is marked with an asterisk.

Fig. 15.

Structure of the pruned NARV model; a minor quadratic term in the Y1 ANF has been omitted from the figure for clarity. As can be seen, the model structure is dominated by self-terms, rather than cross-terms as in the PDM model, and there is no obvious functional interpretation.