SCIENTIFIC REPORTS

OPEN

Automatic Construction of Predictive Neuron Models through Large Scale Assimilation of Electrophysiological Data

Alain Nogaret[1], C. Daniel Meliza[2], Daniel Margoliash[3] & Henry D. I. Abarbanel[4,5]

We report on the construction of neuron models by assimilating electrophysiological data with large-scale constrained nonlinear optimization. The method implements interior point line parameter search to determine parameters from the responses to intracellular current injections of zebra finch HVC neurons. We incorporated these parameters into a nine ionic channel conductance model to obtain completed models which we then use to predict the state of the neuron under arbitrary current stimulation. Each model was validated by successfully predicting the dynamics of the membrane potential induced by 20–50 different current protocols. The dispersion of parameters extracted from different assimilation windows was studied. Differences in constraints from current protocols, stochastic variability in neuron output, and noise behave as a residual temperature which broadens the global minimum of the objective function to an ellipsoid domain whose principal axes follow an exponentially decaying distribution. The maximum likelihood expectation of extracted parameters was found to provide an excellent approximation of the global minimum and yields highly consistent kinetics for both neurons studied. Large scale assimilation absorbs the intrinsic variability of electrophysiological data over wide assimilation windows. It builds models in an automatic manner treating all data as equal quantities and requiring minimal additional insight.

The extraction of information hidden from observation is essential to understanding and modelling complex nonlinear systems[1]. This approach is relevant to the construction of models of biological neurons which relate unobserved ionic transport at the molecular level to real time macroscopic observations of the membrane voltage. While remaining largely inaccessible to experiment, the dynamics of numerous voltage-gated ionic channels acting in parallel through complex nonlinear relationships plays a critical role in shaping the membrane potential[2]. Statistical inference methods can therefore be valuable for extracting the nonlinear parameters controlling ion channel dynamics when coupled with accurate ionic conductance models of the neurons. This is part of the larger effort to understand the higher functions of neural systems through simulation[3]. Several data assimilation procedures including random parameter search[4–8], evolutionary[9] and genetic algorithms[10–15], gradient descent methods[16] and simulated annealing[17] have used semi-empirical Hodgkin-Huxley models[2] to fit linear parameters such as the maximal conductances of ion channels while using tabulated values for the nonlinear parameters such as gate thresholds and time delays[18–20]. Of potential utility would be to identify nonlinear optimization algorithms for data assimilation that would further improve model accuracy by extracting all parameters from experimental data. Such approaches are underpinned by Takens' embedding theorem that states that all information required to constrain the model is contained in the observation of one state variable – the membrane voltage - over a finite time window[21,22]. Two high dimensional nonlinear parameter search algorithms have been tested on conductance models. Vavoulis et al.[23] have estimated *Lymnaea* motoneuron parameters by performing time series analysis with Kitagawa's self-organizing state space approach[24,25]. Meliza et al.[26] have built models of neurons from the zebra finch forebrain nucleus HVC using interior point optimization[27,28].

[1]Department of Physics, University of Bath, Bath BA2 7AY, UK. [2]Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA. [3]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA. [4]Department of Physics, University of California San Diego, La Jolla, CA 92093, USA. [5]Scripps Institution for Oceanography, Marine Physical Laboratory, La Jolla, CA 92093, USA. Correspondence and requests for materials should be addressed to A.N. (email: A.R.Nogaret@bath.ac.uk) or H.D.I.A. (email: habarbanel@ucsd.edu)

Here we demonstrate a large-scale nonlinear optimization method for building complex neuron models. We describe methodology that allowed us to transfer information from observations of the membrane voltage to biologically relevant models. Meliza *et al.*[26] previously studied a large population of neurons using a benchmark electrophysiological protocol which was required to categorize different classes of neurons. The present work instead investigates the responses of individual neurons to a wide range of current protocols. The wide dynamic range covered by these protocols allows us to validate the constructed models both in terms of their predictive power and the plausibility of the extracted parameters.

The method implements Interior Point OPTimization line parameter search (IPOPT)[27,29]. We used this approach to assimilate large sets of electrophysiological data from two HVC neurons and inferred the 71 parameters of a multichannel conductance model. Following assimilation, the parameter solutions were incorporated into the conductance model to obtain a completed model for each neuron. Completed models were then used to predict the state of HVC neurons by forward integration of time series current data. IPOPT minimizes a least square function measuring misfit between observations of the membrane voltage and the corresponding state variable in the conductance model. Minimization is subject to both inequality constraints that specify the search interval of model parameters and equality constraints that specify the rate of change of state variables prescribed by the nonlinear conductance model. Interior point optimization replaces inequality constraints with logarithmic barriers bounding the search domain. These initially provide a convex surface ensuring smooth convergence of the parameter search. The accuracy of the solution is improved through subsequent iterations whilst the barrier decreases. The IPOPT algorithm implements a line parameter search filter which, under mild assumptions on the model, is designed to prune spurious solutions to retain the true solution at the global minimum of the objective function[30].

The completed models were validated by successfully predicting the outcome of experiments that subjected individual neurons to 20–50 current injection protocols some of which had complex dynamics that included both steps and chaotic currents. By comparing the experimental membrane voltages to predictions, we identified the criteria that the assimilation current protocol needs to fulfil in order to effectively constrain the model parameters. The inferred initial state of the neuron was correctly found to be the steady state as expected for an isolated neuron with no prior external stimulation. The biological relevance and uniqueness of parameter solutions rests on the accuracy with which the model describes microscopic gating mechanisms. We addressed this by reformulating our conductance model to comply with the dual needs of biological accuracy and mathematical stability required by constrained optimization. We performed a statistical study of the parameters extracted from different epochs. This identified the recovery time constants as the parameters with the highest degree of functional overlap ("sloppiness")[31,32]. Extracted parameters were found to lie on an ellipsoid surface whose principal axes have an exponentially decaying distribution. A much improved estimate of the global minimum was obtained by calculating the maximum likelihood expectation (MLE) of extracted parameters. The method inferred similar values for the kinetics of the two neurons studied.

Our large scale assimilation method offers many advantages over existing methods. In our approach, all data are assimilated as equal quantities. Completed models were constructed automatically from a generic conductance model that had been previously identified based on biological experiments that relied on pharmacological manipulations to positively identify relevant classes of ionic currents[33]. Automatically generated predictions were in excellent agreement with experiment over virtually all epochs. Biological intuition is minimal and limited to the choice of the parameter search intervals. The assimilation of large scale data absorbs the intrinsic fluctuations of biological neurons to make predictions with accuracy sufficient to evidence trial-to-trial fluctuations in neuron behaviour.

Results

Nonlinear optimization framework. A standard brain slice preparation that included HVC was used for *in-vitro* intracellular recordings of HVC neurons. Neurons were stimulated with a wide range of intracellular injections of current protocols in a series of measurement epochs e (Supplementary, Fig. S1). We used these data to construct a database of epochs recording the membrane voltage $V_{data}^{(e)}(t)$ and current stimulation $I_{data}^{(e)}(t)$. A data assimilation window of duration T was selected from the epoch most suitable for assimilation (Epoch 0). The assimilation window contained $N+1$ observations of the membrane voltage at times $t_i = iT/N$, $i = 0, 1 \ldots N$. Optimization was performed by minimizing a cost function which measures the discrepancies between the experimental voltage and the model output V:

$$c(\boldsymbol{x}(0), \boldsymbol{p}) = \frac{1}{2} \sum_{i=0}^{i=N} \{(V_{data}(t_i) - V(t_i, \boldsymbol{x}(0), \boldsymbol{p}))^2 + u(t_i)^2\}. \tag{1}$$

We represent the state of the neuron by a time dependent vector $\boldsymbol{x}(t)$ which has L vector components. The first component tracks the membrane voltage $x_1(t) \equiv V(t)$ and the others $\{x_l(t)\}_{l=2,3\ldots L} \equiv \{m(t), h(t)\}$ the state of activation and inactivation gates of individual ion channels (Table 1). Overall our conductance model incorporates 9 ion channels giving $L = 12$ state variables. The model also has K parameter constants which contain information about the ion channel conductances, gate thresholds, gate time delays and electric parameters of the neuron membrane such as capacitance and reversal potentials. These parameters, listed in Table 2, are stored in the $K = 71$ vector-components of vector $\boldsymbol{p}$. The objective of nonlinear optimization is to find the parameters $\boldsymbol{p}^*$ and initial conditions $\boldsymbol{x}^*(0)$ which minimize the cost function. The function $u(t)$ in Eq. 1 was used to smooth convergence of the parameter search by eliminating the occurrence of positive conditional Lyapunov exponents[34]. $u(t)$ was defined as a control variable of the assimilation procedure which vanishes as the parameter search approaches the global minimum.

The state variables obey the nonlinear rate equations:

ID	Channel	Current density	Nominal conductance[36,37]
NaT	Fast and transient Na$^+$ current	$J_{NaT} = g_{NaT}m^3h(E_{Na} - V)$	$g_{NaT} = 110\,\text{mS.cm}^{-2}$
NaP	Persistent Na$^+$ current	$J_{NaP} = g_{NaP}m(E_{Na} - V)$	$g_{NaP} = 0.064\,\text{mS.cm}^{-2}$
K1	Transient depolarization activated K$^+$ current	$J_{K1} = g_{K1}m^4(E_K - V)$	$g_{K1} = 5\,\text{mS.cm}^{-2}$
K2	Rapidly inactivating K$^+$ current (A current)	$J_{K2} = g_{K2}m^4h(E_K - V)$	$g_{K2} = 12\,\text{mS.cm}^{-2}$
K3	Ca^{2+} activated K$^+$ current	$J_{K3} = g_{K3}m(E_K - V)$	$g_{K3} = 9.1\,\text{mS.cm}^{-2}$
CaL	High threshold Ca^{2+} current	$J_{CaL} = \rho m^2 J_{Ca}$	—
CaT	Low threshold Ca^{2+} current	$J_{CaT} = m^2 h J_{Ca}$	—
HCN	Hyperpolarization-activated cation current	$J_{HCN} = g_{HCN}h(E_{HCN} - V)$	$g_{HCN} = 0.092\,\text{mS.cm}^{-2}$
Leak	Leakage channels (K & Na)	$J_L = g_L(E_L - V)$	$g_L = 0.066\,\text{mS.cm}^{-2}$

Table 1. Ion channels included in the model. Parameters include the ion conductances g_α, $\alpha \equiv \{NaT, NaP, K1, K2, K3, CaL, CaT, HCN, \text{and } Leak\}$; the sodium and potassium reversal potentials, E_{Na} and E_K; the hyperpolarized-activated cation reversal potential $E_{HCN} = -43\,\text{mV}$[55]. m and h are the state variables of the activation and inactivation gates. J_{Ca} is the Calcium current given by the Goldman-Hodgkin-Katz equation[54].

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{p}). \tag{2}$$

These equalities place L constraints $\boldsymbol{C}(\boldsymbol{X}) = \boldsymbol{0}$ at each point of the assimilation window where $\boldsymbol{X} \equiv \{\boldsymbol{p}, \boldsymbol{x}(0)\}$ are the generalized optimization parameters including the initial conditions. We linearized Eq. 2 at each node t_i using Simpson's method[35] to obtain $L(N+1)$ equality constraints. A second set of constraints was specified by setting the minimum and maximum boundaries of the parameter search interval: $\boldsymbol{p}_L \leq \boldsymbol{p} \leq \boldsymbol{p}_U$. Minimizing the cost function in the presence of both types of constraints was done seeking the extrema of a Lagrangian function $\mathcal{L}(\boldsymbol{X})$ that included both Lagrange multiplier terms for the equality constraints and logarithmic barrier terms for the inequality constraints[27,29]. Logarithmic barriers formulated the inequality constrained optimization problem into an equality constrained convex problem which could be solved by Newton's method. The extrema of the Lagrangian were obtained by solving $\nabla \mathcal{L} = \boldsymbol{0}$ iteratively reducing the barrier height at each iteration. The convex nature of the problem means that the Hessian $\nabla^2 c(\boldsymbol{X})$ and Jacobian $\nabla \boldsymbol{C}(\boldsymbol{X})$ used in Newton's method are smooth at boundaries ensuring that the parameter search converges to the global minimum even for poor choices of starting points in parameter space. Interior point optimization is well suited to neuron model building because it accommodates the very large number of constraints ($>10^6$) needed to assimilate complex electrophysiological data.

Conductance model. We reformulated our conductance model to comply with the stability requirements of the inverse problem while retaining the level of detail provided by the Boltzmann functions used for fitting biological neurons[2,36,37]. The need for the Hessian to be continuous and well-behaved called for rate functions $\boldsymbol{f}$ to be doubly differentiable over the $[-120\,\text{mV}, +50\,\text{mV}]$ range of variation of the membrane voltage. This ruled out the use of functions defined in parts which are convenient fitting functions in electrophysiology[36,37]. In addition, the model had to describe the kinetics with a minimum of parameters to minimize functional overlap between them and warrant uniqueness of the solution.

The first equation in Eq. 2 is the current conservation equation which governs the rate of change of the membrane voltage:

$$C\frac{dV(t)}{dt} = -J_{NaT} - J_{NaP} - J_{K1} - J_{K2} - J_{K3} - J_{CaL} - J_{CaT} - J_{HCN} - J_{Leak} + I_{data}(t)/A, \tag{3}$$

where C is the membrane capacitance per unit area, J_α, $\alpha \equiv \{NaT, NaP, K1, K2, K3, CaL, CaT, HCN, \text{and } Leak\}$ are the current densities of each ion channel, and A is the surface area of the neuron membrane through which the external current is injected. A regularization term $u(t)[V(t) - V_{\text{data}}(t)]$ was added to the right hand side of Eq. 3 to perform assimilation. This term stabilizes convergence of the parameter search by smoothing irregularities in the objective function which are likely to appear when the logarithmic barrier decreases[34]. The ionic currents are chosen based on prior biological knowledge. Each one has a unique mathematical form (Table 1) which facilitates the assignment of parameters and minimizes model degeneracy. An advantage of our data assimilation method is that it does not require biological intuition to assume which ion channel might or might not be present in the experiment because the IPOPT filter automatically assigns a null value to the conductances of the missing channels. A generic conductance model may therefore be used to fit the behavior of different categories of neurons.

The remaining rate equations in the differential system of Eq. 2 describe the first order dynamics of gate variables m and h:

$$\frac{dm(t)}{dt} = \frac{m_\infty(V) - m(t)}{\tau_m(V)}, \tag{4}$$

Ion	Parameter		Neuron 1		Neuron 2	
			p_L, p_U	p^*	p_L, p_U	p^*
	C	μF.cm^{-2}	0.3, 4.0	3.80	0.4, 4.0	0.447
	E_{Na}	mV	42, 50	45.57	40, 54	41.55
	E_K	mV	$-90, -80$	-88.29	$-95, -60$	-95
	E_{Leak}	mV	$-110, -40$	-64.32	$-120, -40$	-67.2
	A	μm^2	$20^2 - 50^2$	38.25^2	$20^2 - 50^2$	24.8^2
NaT	g_{NaT}	mS.cm^{-2}	100, 120	120	70, 130	70
m	V_m	mV	$-49, -27$	-37.85	$-52, -1$	-1.636
	dV_m	mV	5, 32	9.16	3, 32	32.0
	dV_{tm}	mV	5, 45	5.0	3, 45	3
	t_{0m}	ms	0.02, 0.7	0.388	0.001, 0.7	0.00647
	ε_m	ms	0.012, 7	0.012	0.001, 7	0.06239
h	V_h	mV	$-79, -39$	-65.70	$-93, -9$	-29.35
	dV_h	mV	$-35, -5$	-9.76	$-35, -5$	-5
	dV_{th}	mV	4, 43	41.67	5, 47	37.89
	t_{0h}	ms	0.02, 90	0.02	0.02, 180	0.02
	ε_h	ms	1, 470	4.552	1, 970	13.75
NaP	g_{NaP}	mS.cm^{-2}	0, 20	0.0062	0, 19	0.9693
m	V_m	mV	$-69, -29$	-57.37	$-69, -19$	-49.80
	dV_m	mV	5, 32	9.16	3, 45	38.42
	dV_{tm}	mV	5, 45	5.0	3, 45	45.0
	t_{0m}	ms	0.02, 0.7	0.388	0.001, 7	1.213
	ε_m	ms	0.012, 7	0.012	0.001, 7	0.001
K1	g_{K1}	mS.cm^{-2}	0, 80	0.606	0, 130	130
m	V_m	mV	$-69, -21$	-21.0	$-39, 25$	12.25
	dV_m	mV	5, 34	7.561	3, 34	4.860
	dV_{tm}	mV	5, 34	20.11	5, 34	5.0
	t_{0m}	ms	0.01, 5.4	0.34	0.01, 5.4	0.4557
	ε_m	ms	0.002, 23	8.556	0.02, 23	10.04
K2	g_{K2}	mS.cm^{-2}	0, 80	0.430	0, 80	1.128
m	V_m	mV	$-90, -21$	-42.71	$-90, -17$	-83.03
	dV_m	mV	5, 48	10.44	5, 48	48.0
	dV_{tm}	mV	5, 48	7.772	5, 48	27.69
	t_{0m}	ms	0.02, 20	1.889	0.02, 20	4.814
	ε_m	ms	0.5, 23	23.0	0.5, 143	94.81
h	V_h	mV	$-90, -35$	-71.36	$-90, -25$	-52.68
	dV_h	mV	$-39, -5$	-7.49	$-39, -3$	-11.05
	dV_{th}	mV	$-39, -5$	38.99	$-39, -5$	-25.55
	t_{0h}	ms	0.5, 310	189.7	0.5, 310	15.75
	ε_h	ms	0.910	0.5	0.5, 1910	583.4
	δ_h	mV	0, 990	24.35	0, 1990	0
K3	g_{K3}	mS.cm^{-2}	0, 12	0	0, 6	0.5847
m	V_m	mV	$-25, 30$	-24.98	$-85, 10$	$-25, 06$
	dV_m	mV	10, 65	65	10, 65	53.92
	dV_{tm}	mV	10, 70	70	10, 70	17.87
	t_{0m}	ms	0.01, 55	54.97	0.01, 75	4.066
	ε_m	ms	0.1, 455	454.9	0.1, 1955	1955
CaL	g_{out}	mS.cm^{-2}	0.01, 9	0.119	0.01, 9	0.295
	ρ		0.01, 100	0.514	0.01, 100	13.877
m	V_m	mV	$-56, -8$	-56.0	$-69, -8$	-14.21
	dV_m	mV	5, 49	20.49	5, 49	49.0
	dV_{tm}	mV	5, 55	12.95	5, 55	21.295
	t_{0m}	ms	0.09, 43.6	32.17	0.09, 73.6	22.20
	ε_m	ms	0.2, 295	295	0.2, 1895	1895
CaT	V_m	mV	$-80, -35$	-78.8	$-80, -25$	-63.32
m	dV_m	mV	5, 39	39.0	5, 59	59.0
	dV_{tm}	mV	10, 57	53.63	10, 57	10.0

Continued

Ion	Parameter		Neuron 1		Neuron 2	
			p_L, p_U	p^*	p_L, p_U	p^*
	t_{0m}	ms	0.02, 0.9	0.9	0.02, 0.9	0.02
	ε_m	ms	0.5, 97	1.596	0.5, 97	60.63
h	V_h	mV	−90, −55	−57.04	−85, 25	−72.90
	dV_h	mV	−34, −5	−5.0	−44, −3	−29.64
	dV_{t1}	mV	3, 55	55.0	3, 95	43.51
	dV_{t2}	ms	3, 55	55.0	5, 95	95
	t_{0h}	ms	5, 190	50.05	1, 190	11.33
	ε_h	mV	0.5, 7000	539.0	0.5, 1600	1600
HCN	g_{HCN}	mS.cm^{-2}	0, 10	0.013	0, 6	0.2608
h	V_h	mV	−90, −40	−75.50	−90, −40	−66.92
	dV_h	mV	−30, −5	−5.0	−30, −3	−8.142
	dV_{th}	mV	5, 40	40.0	3, 50	4.403
	t_{0h}	ms	0.1, 500	161, 8	0.1, 500	74.741
	ε_h	mV	0.1, 5000	0.1	0.1, 500	33.96
Leak	g_L	mS.cm^{-2}	0.01, 0.6	0.0794	0.01, 0.6	0.6

Table 2. Parameters of completed neuron models extracted from Epoch 0.

$$\frac{dh(t)}{dt} = \frac{h_\infty(V) - h(t)}{\tau_h(V)}, \tag{5}$$

for all 9 ion channels. The steady state activation curves of gate variables and time delays were:

$$m_\infty(V) = \frac{1}{2}\left(1 + \tanh\frac{V - V_m}{dV_m}\right), \ \tau_m(V) = t_{0m} + \varepsilon_m\left(1 - \tanh^2\frac{V - V_m}{dV_{tm}}\right), \tag{6}$$

$$h_\infty(V) = \frac{1}{2}\left(1 + \tanh\frac{V - V_h}{dV_h}\right), \ \tau_h(V) = t_{0h} + \varepsilon_h\left(1 - \tanh^2\frac{V - V_h}{dV_{th}}\right), \tag{7}$$

where V_m and V_h are the gate activation and inactivation thresholds, dV_m and dV_h give the slope of the transition from open to closed state, t_{0m} (resp. t_{0h}) is the recovery time of the activation (resp. inactivation) variable in the closed state while $t_{0m} + \varepsilon_m$ (resp. $t_{0h} + \varepsilon_h$) is the recovery time in the fully open state. The inactivation kinetics of the $K2$ and CaT currents has two distinct recovery rates above and below a transition voltage which require different mathematical descriptions[36,37]. Below $V_h + \delta_h$, the recovery time of $K2$ follows the bell-shaped dependence of Eq. 7. Above the threshold, it becomes independent of the membrane voltage[36]. We describe this two-part behavior with the following equation valid for any membrane voltage:

$$\tau_h(V) = t_{0h} + \varepsilon_h\left(1 - \tanh^2\frac{\delta_h}{dV_{th}} + \frac{1}{2}(1 - \tanh(1 - V_{th} - \delta_h))\left(\tanh^2\frac{\delta_h}{dV_{th}} - \tanh^2\frac{V - V_{th}}{dV_{th}}\right)\right). \tag{8}$$

The inactivation kinetics of the CaT channel is similarly controlled by two recovery times giving a bi-exponential dependence[36]. The two part exponential is combined into a single equation:

$$\tau_h(V) = t_{0h} + \varepsilon_h \frac{\left(1 + \tanh\left(\frac{V - V_h}{dV_{t1}}\right)\right)\left(1 - \tanh\left(\frac{V - V_h}{dV_{t2}}\right)\right)}{1 + \tanh\left(\frac{V - V_h}{dV_{t1}}\right)\tanh\left(\frac{V - V_h}{dV_{t2}}\right)}$$
$$\times\left((1 - \tanh(V - V_h))\tanh\left(\left(\frac{1}{dV_{t1}} + \frac{1}{dV_{t2}}\right)(V - V_h)\right)\right), \tag{9}$$

where parameters dV_{1t} and dV_{2t} define the recovery times above and below the voltage threshold V_h. Our model equations Eqs 6–9 were validated by fitting successfully the experimental activation functions of thalamocortical neurons (Supplementary Fig. S2).

Model predictions. The assimilated neuron models were primarily validated by testing their ability to predict multiple epochs implementing a wide range of current protocols (see methods). To this end, we constructed the fully automated assimilation-prediction procedure depicted in Fig. 1. In the first stage, the inputs of IPOPT are the electrophysiological recordings $V_{data}^{(0)}(t_i)$ and $I_{data}^{(0)}(t_i)$, $i = 0, 1 \ldots N$ chosen to assimilate data from Epoch 0, and the boundaries of the parameter search intervals, $\boldsymbol{p}_L$ and $\boldsymbol{p}_U$. IPOPT outputs the state vector that minimizes the objective function at each point of the assimilation window $\boldsymbol{x}^*(t_i)$, $i = 0, 1 \ldots N$ and the parameter vector

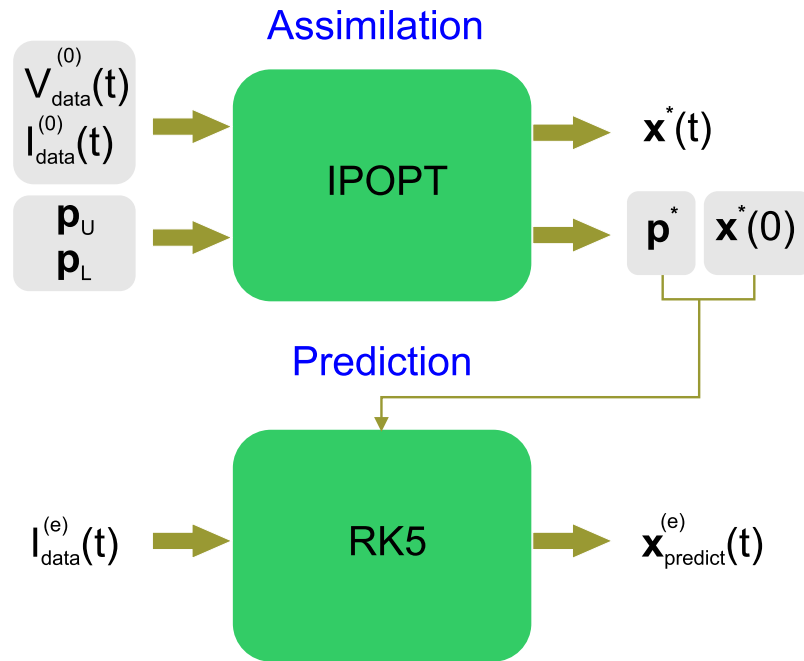**Figure 1. Assimilation-prediction method.** The injected current waveform $I_{data}^{(0)}(t)$ and resultant time series membrane voltage $V_{data}^{(0)}(t)$ of the epoch to assimilate (Epoch 0) are input into the nonlinear optimization filter (IPOPT). These provide the equality constraints. The user specifies the inequality constraints by choosing the upper and lower boundaries of the parameter search intervals, $\mathbf{p}_L$ and $\mathbf{p}_U$. IPOPT outputs the state variables $\mathbf{x}^*(t)$ solution of the minimization problem at each point of the assimilation window, together with the parameter solutions $\mathbf{p}^*$. The extracted parameters $\mathbf{p}^*$ are inserted in the model equations to construct the *completed* model. The state of the neuron $\mathbf{x}_{predict}(t)$ is predicted by integrating the current protocol $I_{data}^{(e)}(t)$ forward from initial conditions $\mathbf{x}^*(0)$ using a fifth order, adaptive step size, Runge-Kutta solver (RK5). The model is validated by comparing the predicted membrane voltage $V_{predict}^{(e)}(t)$ with the voltage $V_{data}^{(e)}(t)$ recorded in Epoch e.

solution $\mathbf{p}^*$. In the second stage, the $\mathbf{p}^*$ were inserted into the model equations Eqs 6–9 to obtain the *completed model*. This model was then used to predict the state of this neuron $\mathbf{x}_{predict}^{(e)}(t)$ by forward integrating the experimental current protocol of another epoch $I_{data}^{(e)}(t)$. By default, the initial conditions $\mathbf{x}(0)$ at the start of integration were obtained from data assimilation (see below). Although the system of equations in Eq. 2 is not believed to be chaotic, the multiplicity of recovery times arising from 9 ion channels induces system stiffness and reduces tolerance to integration error. It was therefore necessary to implement forward integration using adaptive step size fifth order Runge-Kutta (RK5) to achieve the required level of accuracy[38].

Figure 2 shows the two reference epochs which we use to construct the completed models of two representative HVC neurons: a putative RA-projecting neuron (N1) and a putative X-projecting neuron (N2). The current protocol (blue line) induced the oscillations observed in the membrane voltage (black line). Data assimilation was performed over 1600 ms long time interval (N1) and 900 ms (N2) using $N = 80,000$ and $N = 90,000$ mesh points respectively. The mesh size T/N was chosen to sample potential spikes with ≈ 100 data points each. The width of the data assimilation window was chosen as a tradeoff between the need to incorporate a statistically meaningful number of spikes and the need to minimize numerical error that accrues when handling larger Jacobian and Hessian matrices. With a constant mesh size, the optimum number of data points was empirically found to be $N \approx 100,000$.

The state variables $\mathbf{x}^*(t_i)$, $i = 0, 1 \ldots N$ solution of the constrained optimization include the membrane voltage $V^*(t_i)$ (green line, Fig. 2) and the 11 gate variables (Supplementary Figs S5 and S6). The parameter solutions $\mathbf{p}^*$ are listed in Table 2. The completed models of both N1 and N2 are found to synchronize well to the experimental data. The fitting error was maximum at the site of voltage spikes where $|(V^*(t_i) - V_{data}(t_i))/V_{data}(t_i)| \leq 6\%$ (Supplementary Fig. S4). Elsewhere, the fit of subthreshold oscillations produced less than 1% error.

The state of each neuron was predicted beyond the end of the assimilation window by integrating the completed model forward (red curves, Fig. 2). We used the state variables $\mathbf{x}^*(T)$ generated by IPOPT at the end of the assimilation window as initial conditions for forward integration. The predicted oscillations of the membrane potential are identical whether forward integration starts from the beginning of the assimilation window or from the end. The predicted timings of voltage spikes, spike shape, spike amplitude and sub-threshold oscillations are in remarkable agreement with the experiment. Occasionally current stimulation near the firing threshold causes a missing or added spike. These discrepancies are likely to be due to spontaneous synaptic activity in the slice rather than from model inaccuracy. This can be seen in Supplementary Fig. S7 (an expanded plot of Fig. 2) where three spikes at $t = 1653$, 1657 and 1662 ms (blue arrows) are missing at the centre of a spike burst but are predicted by the model. Because the amplitude of current stimulation increases progressively from 1640 ms to 1680 ms, N2
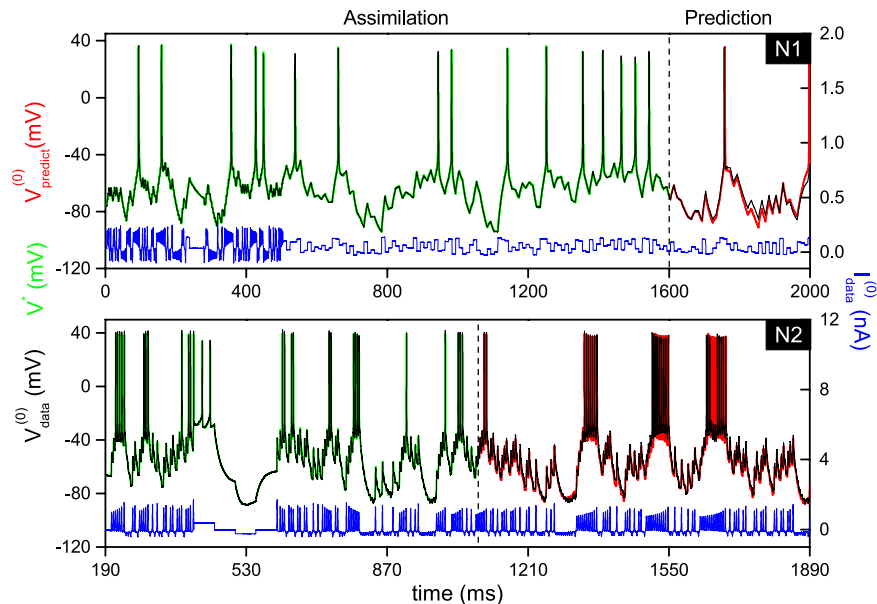
Figure 2. Membrane voltage oscillations used to extract the model parameters of two HVC neurons (Epoch 0). The experimental voltage $V_{data}^{(0)}(t)$ (black line) was recorded under current clamp stimulation by current waveform $I_{data}^{(0)}(t)$ (blue line). N1 is a putative RA-projecting neuron (top panel) and N2 is a X-projecting neuron (bottom panel). The assimilation window used to extract the parameters spans the interval [0–1600 ms] for N1 and [190 ms–1090 ms] for N2. The membrane voltage solution of the constrained optimization problem is $V^*(t)$ (green line). The membrane voltage predicted by integrating the experimental current waveform is $V_{predict}^{(0)}(t)$ (red line). Details of the oscillations of N2 are plotted in supplementary Figure S7.

could be expected to burst without interruption as it does in the preceding bursts starting at 1340 ms and 1520 ms. This observation underlines an important benefit of large scale data assimilation for integrating stochastic fluctuations in neuron output[39]. By choosing long assimilation windows, stable models may be constructed from imperfect data. Forward integration of Eq. 2 also predicted the dynamics of unobserved gate variables (Supplementary Figs S5 and S6) and ionic currents[26].

The completed models are then used to predict the state of a neuron stimulated by different current protocols $I_{data}^{(e)}$ (Epochs $e = 1$–50). The four protocols shown in Fig. 3 test model predictions to current waveforms with different shapes, amplitudes and instantaneous frequencies. Predictions are made from $t = 0$ onwards using initial conditions $\boldsymbol{x}^*(0)$ computed by IPOPT. Panels (a) and (b) show the predicted voltage of N1 in 1 s long snapshots at the beginning and at the end of a 6 s long epoch through which forward integration has run continuously. The good agreement between the predicted voltage (red line) and the experimental voltage (black line) demonstrates the stability of forward integration over long time intervals. Another factor explaining the good match is the slower rate of oscillations of the stimulating current which is well within the bandwidth of the assimilation protocol (Supplementary Fig. S3). The current stimulation in panels (a),(b) therefore tests time constants of the model which are well constrained by the assimilation protocol. The current protocol of panel (c), in contrast, incorporates oscillations twice as fast as in the assimilation protocol. The predicted voltage occasionally exhibits extra spikes caused by near threshold current oscillations at 90 ms, 430 ms, 450 ms and 950 ms. Panel (d) plots the response to current steps wider than those used in the assimilation protocol (32 ms instead of 10 ms). Although predictions remain good, discrepancies begin to emerge in the response to longer current steps. This can be seen near the end of Epoch 4 where two consecutive current steps effectively form a 64 ms wide pulse. This observation corroborates the fact that tonic spiking elicited by near threshold direct currents is notoriously difficult to reproduce due to stochastic resonance[39] and oscillations of the subthreshold membrane potential[40].

We turn next to neuron N2. Both current waveforms in Fig. 3(e,f) are more intricate than the one used in the assimilation step as they mix large amplitude current oscillations, steps, and square pulses modulated by chaotic oscillations. The current oscillations in panel (e) are slower than in the assimilation window (11 ms interspike interval vs 5 ms) whereas they are faster in panel (f) (3 ms interspike interval vs 5 ms). Panels (g) and (h) focus on slowly varying currents. A good agreement is obtained between the predicted and the observed voltages in panel (e) which are driven by relatively slow-varying currents. The response to a rapidly varying current at the start of panel (f) shows several spikes missing from the predicted curve. Prediction here, is complicated by the very short (3 ms) time interval between consecutive current pulses which approaches the width of a voltage spike (~2 ms). The instantaneous frequency of these current oscillations (330 Hz) is above the 200 Hz cut-off frequency of the assimilation protocol (Supplementary Fig. S3). These observations underline the need for wideband assimilation protocols to constraint all time constants of the model.

We then tested the model ability to predict the excitatory response hence the current threshold of a neuron. The experiment was performed by applying a depolarizing current step of 200 ms duration followed by a hyperpolarizing step (Fig. 4). The amplitude of the depolarizing current was increased through the threshold of the
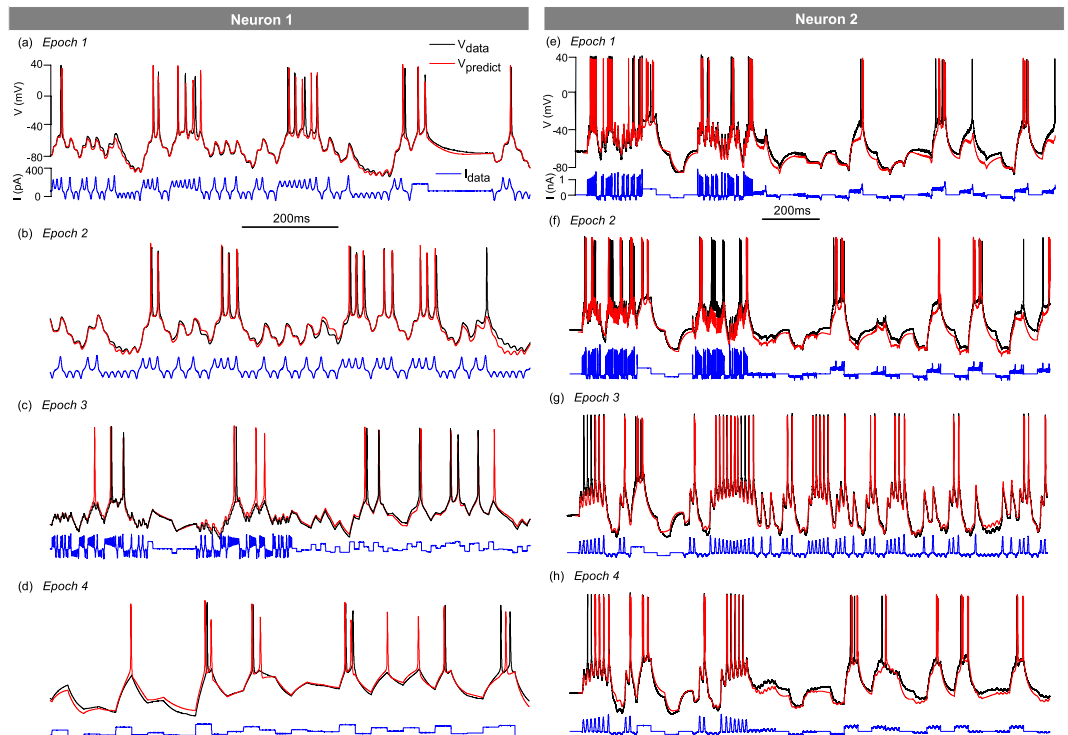
Figure 3. Validation of completed models by comparing predicted membrane voltages (red lines) with experimentally measured voltages (black lines). The current protocols in Epochs 1–4 (blue lines) include a broad range of waveforms and timescales which are distinct from those used to assimilate data. Each current protocol was integrated from the origin onwards to obtain the predicted voltage (red line). The same set of parameters was used to predict all Epochs of N1 (resp. N2).

neuron in steps of 20 pA (Panels (a)–(e)). Predictions were made by forward integrating the experimental current protocols of panels (a)–(e) with the completed model of N1. The predicted output in each case is plotted in panels (f)–(j). The model describes the main features of the excitatory response namely, the increase in firing frequency with increasing current stimulation, the correct number spikes per burst (± 1 spike), the decay in spike amplitude and widening of inter-spike intervals. The model exaggerates the rebound of the membrane voltage after the release of hyperpolarization. These results show that a single assimilation protocol (Fig. 2) is sufficient to evaluate the firing threshold to a good degree of accuracy: 50 ± 5 pA. Spike bursts evoked by long current steps are difficult to validate through direct comparison with experiment in the manner of Figs 2 and 3 because of the poor reproducibility of real neurons under tonic stimulation[39,40].

Validation of initial conditions. The estimation of initial conditions is essential to predicting the state of the neuron, as erroneous starting values often lead to dramatically different behavior over time scales exceeding the relaxation times of the system. The initial state of the neuron $x(0)$ may obtained as a solution of the minimization problem (Eq. 1). Network activity is relatively suppressed in these brain slice recordings, hence it may be reasonably assumed that the initial state of the HVC neuron is the steady state implying $\dot{x}(0) = 0$. We inserted this condition in Eq. 2 which we solved to obtain the state variables of the neuron in the steady state $x_{steady}(0)$. Figure 5 compares the neuron output predicted using initial conditions generated by IPOPT (green and blue lines) and steady state initial conditions (red lines). Results are calculated over a 2 s long protocol (top panel) from which we single out two shorter time intervals at the beginning (panel A) and at the end (panel B) for discussion. Steady state initial conditions (red line) give near perfect predictions in panel A. Initial conditions obtained from assimilation also produce good predictions (to within a few additional spikes). At times larger than the relaxation times of the system, all differences arising from initial conditions vanish and predictions become identical (panel B). This experiment therefore demonstrates that the actual initial state of the neuron is the steady state showing that the neuron is effectively isolated from its environment. We have also verified that the assimilation procedure correctly infers physically meaningful initial conditions.

Bayesian analysis of extracted parameters. Ideally, the assimilation of different epochs of the same neuron ought to yield identical sets of parameters. In practice, differences in constraints from different current protocols, stochastic variability in neuron output, noise and ill-posedness of the inverse problem introduce uncertainty in the parameter field. We make constructive use of this randomness to obtain additional information about the model. We have extracted R sets of parameters $p_1^\star, p_2^\star \ldots p_R^\star$ from R different epochs to construct a sample of possible values taken by random vector P. Gaussian uncertainty in the probability density implies the solutions of
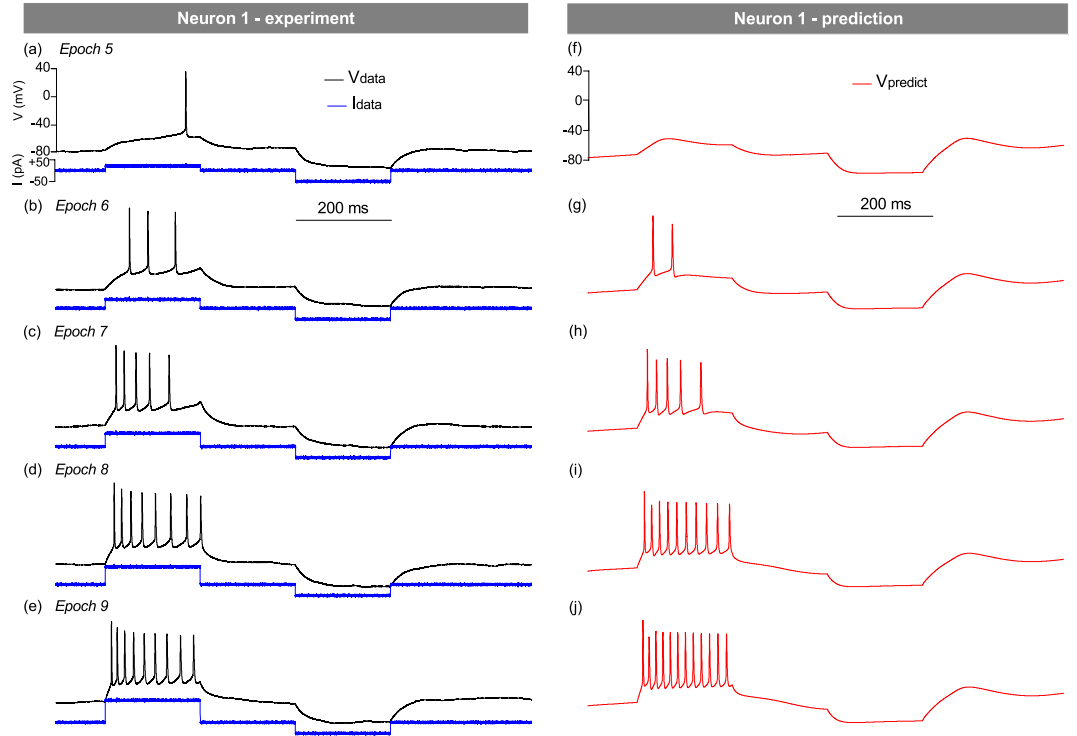
Figure 4. Experimental and theoretical voltage output by Neuron 1 under stimulation by longer current steps. A depolarizing current step is applied between 100 ms and 300 ms. Its amplitude increases from 50 pA (Epoch 5) to 130 pA (Epoch 9) in steps of 20 pA. A constant hyperpolarizing current of -80 pA is applied between 500 ms and 700 ms in all epochs. The predicted voltage (right column) is obtained by integrating the 5 current protocols with the completed model of N1.

the inverse problem are also Gaussian. This assumes that experimental error is sufficiently small for the forward solutions to cover a small region of parameter space. As a result, the random vector $\boldsymbol{P}$ follows a normal distribution centered on maximum likelihood expectation $\boldsymbol{P}^{\star}$ and characterized by its covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{p}^{\star}}$[41,42]:

$$N(\boldsymbol{P}) = \frac{1}{\sqrt{2\pi|\Sigma_{\boldsymbol{p}^{\star}}|}} \exp\left(-\frac{1}{2}(\boldsymbol{P}-\boldsymbol{P}^{\star})^{T}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{p}^{\star}}^{-1}(\boldsymbol{P}-\boldsymbol{P}^{\star})\right). \tag{10}$$

A Taylor expansion of the objective function Eq. 1 with respect to small changes in parameters about the minimum $\boldsymbol{p}^{\star}$ yields the data misfit:

$$\delta c = \frac{1}{2}(\boldsymbol{p}-\boldsymbol{p}^{\star})^{T}\hat{\boldsymbol{H}}(\boldsymbol{p}-\boldsymbol{p}^{\star}), \tag{11}$$

where $\hat{\boldsymbol{H}}$ is the Hessian matrix of the data misfit term δc. Identification of δc with the argument of the exponential in Eq. 10 yields $\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{p}^{\star}}^{-1}$. Eq. 11 shows that the surface of constant misfit is a K-dimensional ellipsoid in parameter space. The lengths of its principal axes determine the degree of correlation (or sloppiness) between parameters[31]. These are obtained by calculating the square roots of the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{p}^{\star}}$: $\{\varepsilon_k\}$, $k = 1, 2 \ldots K$. The amplitude of eigenvalues determines the degree of functional overlap between model parameters.

We have assimilated the data from $R = 84$ time windows for N1 (Supplementary Fig. S8) and likewise for N2 (Supplementary Fig. S9) to obtain forward parameter solutions $\{\boldsymbol{p}_{r}^{\star}\}$, $r = 1, 2 \ldots 84$. This number was chosen because the number forward solutions needed to compute the covariance matrix of the posterior probability density $N(\boldsymbol{P})$ had to be greater than the number of parameters ($R > L$). We then constructed the covariance matrix of dimensionless parameters $\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{p}}^{\star}}$ by normalizing individual parameters with respect to their search interval $\boldsymbol{p}_{U} - \boldsymbol{p}_{L}$. Normalization allows computing meaningful eigenvalues from parameters expressed in different units. The normalized parameters were calculated as: $\tilde{p}_{r,l} = p_{r,l}/(p_{U,l} - p_{L,l})$, $l = 1, 2 \ldots L$, $r = 1, 2 \ldots R$. The dimensionless covariance matrix was calculated as:

$$s_{k,k'} = \frac{1}{R-1}\sum_{r=1}^{R}\left(\tilde{p}_{r,k} - \overline{\tilde{p}}_{k}\right)\left(\tilde{p}_{r,k'} - \overline{\tilde{p}}_{k'}\right) \qquad k, k' = 1, 2 \ldots K \tag{12}$$

where the $\overline{\tilde{p}}_{k}$ are the mean values of normalised parameters averaged over the statistical sample. The covariance matrix of N1 is mapped in Fig. 6(a). The spectrum of eigenvalues of the N1 and N2 matrices is plotted in Fig. 6(b).
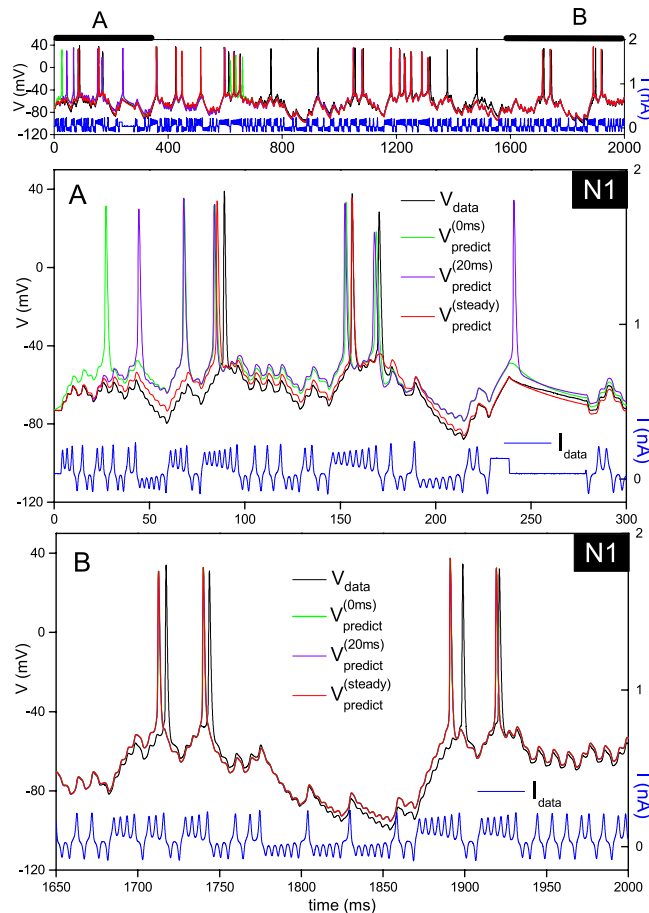
Figure 5. Effect of the choice of initial conditions on model predictions. The initial values of state variables at the beginning of forward integration were obtained by two methods: data assimilation (green and violet lines) and by assuming steady state conditions at the beginning of the epoch (red line). Assimilation of the first 20 ms of data yielded initial conditions at $t = 0$ (green line) and at $t = 20$ ms (violet line). Panel A shows the differences in neuron oscillations induced by different initial conditions over the first 300 ms. These differences however vanish with time. Panel B shows that all predictions become identical near the end of the epoch, irrespective of initial conditions.

Eigenvalues decay exponentially for both N1 and N2 (Supplementary Fig. S10). The majority of these, 57/71, have amplitude less than 10% of the largest eigenvalue which indicates that most parameters are well constrained. To identify the sloppy parameters associated with the larger eigenvalues, we conducted a side by side examination of the covariance matrices of N1 and N2 (Supplementary Fig. S11). The first observation is that finite off-diagonal matrix elements are not random but align in rows and columns which are the same in N1 and N2. Closer examination reveals that sloppy parameters correspond to the recovery times τ_m and τ_h of the *NaT, K2, K3, CaL, CaT* channels. The best constrained parameters are the voltage thresholds V_m and V_h. This can be seen in Fig. 6(c) which compares the standard deviations of voltage thresholds and gate recovery times. For all ionic gates without exception, the gate recovery time has a larger standard deviation than the voltage threshold. This result may be explained by the choice of more complex equations for recovery times which contain more adjustment parameters that the equations of activation curves.

Experimental error has the same effect as a residual temperature T which places a lower limit on the free energy of the system (the cost function c). This temperature prevents cooling the system to zero to reach the global minimum. The best that can be hoped is that direct parameter search arrives on the surface of the ellipsoid $\delta c = kT$ where $k = 2(N+1)k_BR\Delta f$ is the signal noise entropy, k_B Boltzmann's constant, R the resistance of the neuron and Δf the noise bandwidth. It follows that the random vector $\boldsymbol{P}$ maps the ellipsoid surface $\delta c = kT$ which is centered on its MLE. Assuming that T is not so large that the second order expansion Eq. 11 evaluates δc with sufficient accuracy, the true global minimum may be obtained by calculating $\boldsymbol{P}^*$. We have done this for the gate voltage thresholds V_m and V_h of N1 and N2 and plotted the results in Fig. 6(d). Remarkably, the MLEs give highly consistent threshold values in N1 and N2. In addition, the activation functions and recovery times calculated using MLEs are also in very good agreement in N1 and N2 (Fig. 6(e)).

Discussion

Our results show that interior point line parameter search successfully addresses several challenges identified as critical to building neuron models[20,32,43]. Our method extracts both linear and nonlinear parameters from
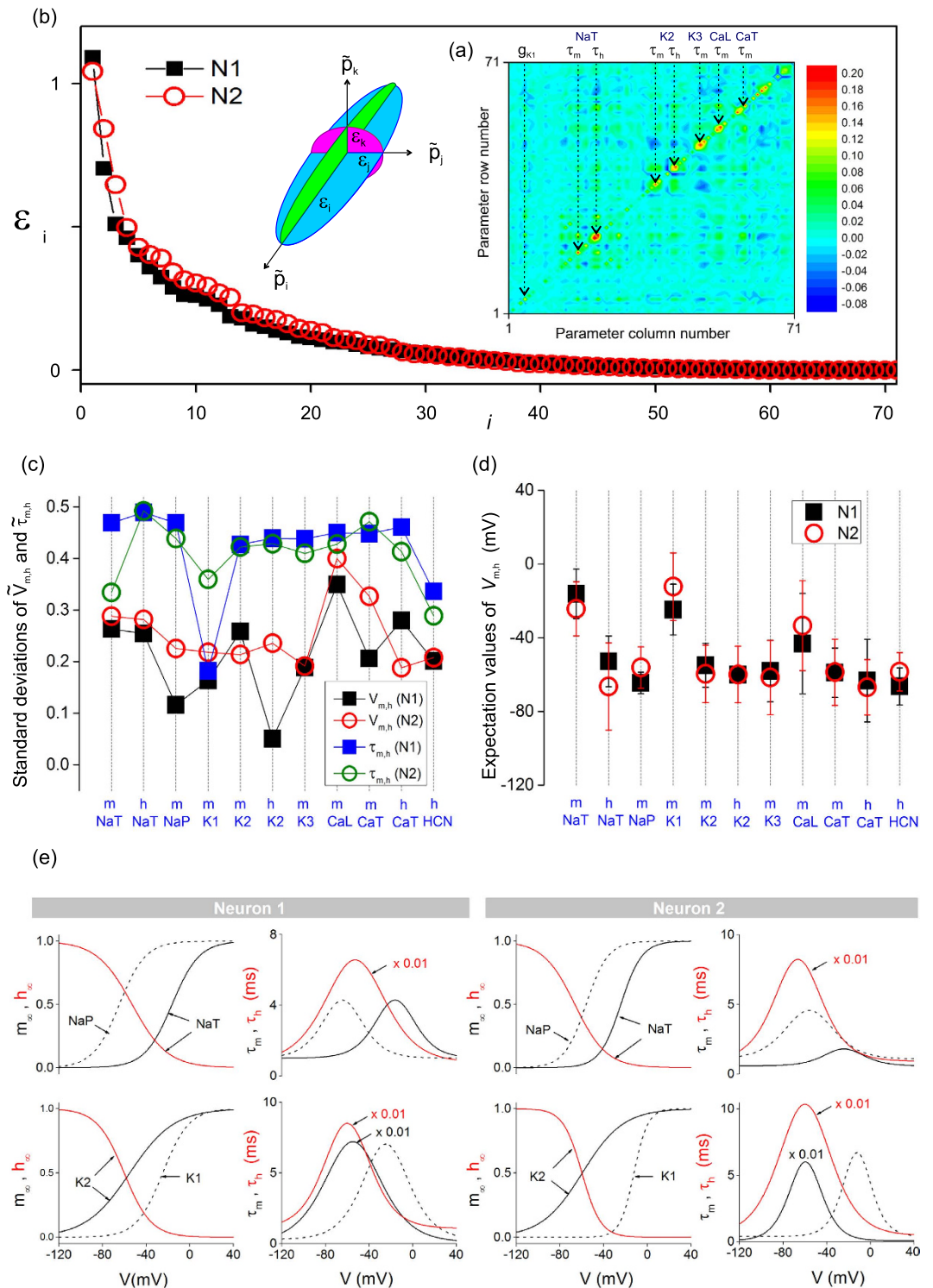
Figure 6. (**a**) Covariance matrix of the random vector $\widetilde{P}$ of N1. The dashed arrows indicate the sloppy parameters with the largest off-diagonal matrix elements. (**b**) Spectrum of eigenvalues of the covariance matrix of N1 (black squares) and N2 (red circles) which measure the length of the principal axes of the data misfit ellipsoid (inset). (**c**) Standard deviations of activation and inactivation voltage thresholds and recovery times of neuron N1 (square symbols) and N2 (circles). (**d**) Maximum likelihood expectation values of the activation and inactivation thresholds of N1 (dark squares) and N2 (red circles). The error bars are the standard deviations. (**e**) Stationary values (m_∞, h_∞) and time constants (τ_m, τ_h) of the activation and inactivation variables of *NaT*, *NaP*, *K*1 and *K*2 ion currents for neurons N1 and N2. These were calculated using the maximum likelihood expectation of parameters.

11

arbitrarily complex time series data. The large scale nature of the problem allows integrating neuron variability and noise over wide assimilation windows to give highly stable completed models. Data assimilation is a fully automatic process that treats all data points as equally important unlike multi-objective methods that select fitting criteria such as spike rate, spike width, spike height, number of spikes in bursts etc[10,17]. Each one of our completed models made successful predictions of the state of neurons stimulated by a multitude of current waveforms. We found that discrepancies between predicted and experimental voltages arose when the assimilation protocol loosely constrained the shortest recovery time constants of the model, in particular those of the transient sodium channel. This occurred when the bandwidth of the assimilation protocol was narrower than that of the protocol used for testing predictions. This condition is sufficient but not necessary as there are other factors to overcome, such as noise, to reach the global minimum of the objective function. Recovery times exhibited increased sloppiness relative to other parameters due to the comparatively larger number of adjustment parameters needed to describe complex gate dynamics.

Toth *et al.*[35] have applied the data assimilation method to model data. They have shown that the solution of well-posed inverse problems involving Hodgkin-Huxley models is single valued. These so called "twin experiments" used the same conductance model to compute the membrane voltage and infer the original model parameters back from it. The challenge of assimilating data from real neurons is that their precise model is unknown. This lack of knowledge was mitigated by analyzing a model of HVC neurons that incorporated ionic currents whose identity had been confirmed previously in biological experiments[33]. Using a model incorporating current types not present in the actual neurons is not problematic since IPOPT assigns near zero values to their ionic conductances. A model lacking an ionic current type expressed in the biological neurons is a more problematic condition. Therefore the possibility of specifying models with numerous ionic currents is an advantageous approach to address the incompleteness of biological knowledge. From a mathematical point of view, the lack of knowledge on the exact equations of the biological neuron makes the problem ill-posed. Simulated annealing experiments have indicated the formation of additional local minima on the free energy surface of conductance models[44]. This possibility was considered in our analysis. However we do not believe local minima to play a role here because the widely different $V_{data}(t)$ observed in N1 and N2 would have produced different free energy surfaces and local minima at different locations in N1 and N2 (see Eq. 1). Parameters having converged to local minima would therefore have produced different covariance matrices and eigenvalue spectra for N1 and N2 which is not what is observed in Supplementary Fig. S11.

Once forward solutions are ascertained to have converged near the global minimum, their maximum likelihood expectation will always provide an even better approximation of the global minimum (Supplementary Fig. S12). The MLE has the additional advantage of cancelling parameter sloppiness. The direct calculation of covariance matrices however carries a high computational cost. A minimum of 71 sets of parameters have to be extracted at a cost of ~50 hours of workstation time per set. Algorithms based on Bayesian uncertainty quantification[41] may provide an alternative path to estimate these matrices.

The choice of the parameter search interval generally is not critical to convergence as the parameter search generates predictive models even when unrealistically large intervals are used. Setting a too tight parameter range or unreasonable boundaries however is not allowed as the model will not fit the data. Conversely, increasing the 71-dimensional search volume beyond reason, needlessly increases computation time. Our parameter search intervals in Table 2 were chosen to encompass the widest range of biologically plausible values which we further incremented by a safety margin.

Our work makes a number of assumptions. Ligand-gated ion channels are omitted. Although the assimilation method may easily be extended to describe multiple compartments, a single compartment neuron model is found to be adequate to describe currents injected in the soma. The relevant compartment is the axon hillock which leads to spike initiation, has a higher density of Na channels, and is weakly coupled to dendrites. Our nonlinear parameter search is generally very robust having successfully constructed models from 22 out of 26 HVC neurons[26]. In summary, large scale data assimilation combined with Bayesian inference is a very effective method for extracting information on microscopic processes and building the accurate neuron models needed to simulate the higher functions of networks.

Methods

Electrophysiology.   Songbird neurons have a complex system of ion channels[45–48] which make them appropriate for testing multichannel conductance models. Wholecell recordings were made in current clamp mode from HVC brain slices prepared from adult male zebra finches (Taeniopugia guttata). Current clamp measurements were made on male zebra finch HVC neurons, and recordings from 26 neurons were selected for further analysis[26]. Out of these we tentatively identified RA-projecting neurons, X-projecting neurons and interneurons which are broad classes of neurons differing in anatomy[49,50], physiological properties[33,51] and presumably circuit function[52]. Neuron N1 was classified as a RA-projecting neuron based on its lack of adaptation and high firing rate. N2 was identified as a X-projecting neuron based on its lack of adaptation. The current clamp hardware and *in-vitro* solutions are described in Supplementary Fig. S1. Between 20 and 50 different current protocols were applied depending on the duration of the recordings for each neuron. Current protocols were designed with the bandwidth needed to constrain the internal time constants of the neuron (Supplementary, Fig. S3). They were composed of aperiodic oscillations synthesized by the chaotic Lorenz system[53] mixed with positive and negative current steps calibrated in amplitude to induce depolarization, subthreshold oscillations and hyperpolarization. A flat power spectrum was desirable to give equal weight to the assimilation of the different time constants of the model. The chaotic Lorenz oscillator model fulfils these requirements and is simple to use for synthesizing current waveforms. Some current protocols consisted exclusively of current steps of varying magnitude and duration to characterize the excitatory response of neurons. The input impedance of the amplifier ($10^{13}\,\Omega$) was orders of magnitude larger than the leakage resistance of any HVC neuron (10^8–$10^{10}\,\Omega$) therefore leakage through the

measurement system could safely be neglected in Eq. 3. Measurements were taken at a 50 kHz sampling rate over periods lasting 2 s–6 s. The data were eventually linearly interpolated to resolve individual spikes with at least 100 points/spike.

Numerical analysis. The IPOPT software inputs two text files. The first one specifies the model equations Eqs 3–11. The second one contains the parameters boundaries $\boldsymbol{p}_L$ and $\boldsymbol{p}_U$, the size of the assimilation widow N, T and the names of the input files containing the data to assimilate. The Jacobian and Hessian matrices $\boldsymbol{\nabla C}$ and $\nabla^2 c$ are calculated using symbolic differentiation (Sympy) prior to compilation of the executable code. Sparse linear systems are solved by embedding the NAG MAS57 solver in IPOPT. The Goldman-Hodgkin-Katz equation describing calcium dynamics[54] was computed as a Taylor series expansion, retaining the first 25 terms to obtain the required accuracy over the $[-120\,\text{mV}, +50\,\text{mV}]$ range of the membrane voltages. The polynomial was computed efficiently in the Horner form:

$$J_{Ca} = V_T \frac{g_{\text{out}} - g_{\text{in}} \exp\left(\frac{V}{V_T}\right)}{1 + \frac{V}{2V_T}\left(1 + \frac{V}{3V_T}\left(1 + \frac{V}{4V_T}(1 + \dots)\right)\right)}$$

(13)

where thermal voltage $V_T = 13\,\text{mV}$ was set as a constant ($T = 25\,°\text{C}$). The outer calcium conductance g_{out} is a parameter of data assimilation. The inner calcium conductance g_{in} is four orders of magnitude smaller than g_{out} and was set at a constant value of $10^{-4}\,\text{mS.cm}^{-2}$ in the model.

The data files containing the parameter solutions output by IOPT and the current protocol to integrate were read by a custom made C program which performed forward integration by implementing adaptive step fifth order Runge-Kutta method[38]. The regularization term $u(t)[V(t) - V_{\text{data}}(t)]$ was left out of the completed model being integrated forward. The covariance matrix of the random parameter vector was calculated within a purposely written C program. This matrix being symmetric, its eigenvalues were calculated by applying Jacobi transformations to its elements[38].

References

1. Abarbanel, H. D. I. Predicting the future: Completing Models of Observed Complex Systems, Springer-Verlag. June (2013)
2. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117,** 500 (1952).
3. Markram, H. The Blue Brain project. *Nature Rev. Neurosci.* **7,** 153 (2006).
4. Prinz, A. A., Billimoria, C. P. & Marder, E. Alternative to hand-tuning conductance-based models: constriction and analysis of databases of model neurons. *J. Neurophysiol.* **90,** 4015 (2003).
5. Golowasch, J. S. G. M., Abbott, L. F. & Marder, E. Failure of averaging in the construction of a conductance-based neuron model. *J. Neurophysiol.* **87,** 1129 (2002).
6. Goldman, M. S., Golowasch, J., Marder, E. & Abbott, L. F. Global structure, robustness, and modulation of neuronal models. *J. Neurosci* **7,** 1345 (2004).
7. Huys, Q. J. M., Ahrens, M. B. & Paninski, L. Efficient estimation of detailed single-neuron models. *J. Neurophysiol.* **96,** 872 (2006).
8. Baldi, P., Vanier, M. C. & Bower, J. M. On the use of Bayesian methods for evaluating compartmental models. *J. Comput. Neurosci.* **5,** 285 (1998).
9. Eiben, A. E. & Smith, J. E. *Introduction to evolutionary computing* (Springer, Heidelberg, 2003).
10. Druckmann, S., Banitt, Y., Schürmann, F., Markram, H. & Segev, I. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front. NeuroSci.* **1,** 7 (2007).
11. Kobayashi, R., Tsubo, Y. & Shinomoto, S. Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Front. Comput. NeuroSci* **3,** 1 (2009).
12. Marasco, A., Limongiello, A. & Migliore, M. Fast and accurate low-dimensional reduction of biophysically detailed neuron models. *Sci. Rep.* **2,** 928 (2012).
13. Achard, P. & De Schutter, E. Complex parameter landscape for a complex neuron model. *PLoS Comput. Biol.* **2,** e94 (2006).
14. Buhry, L., Pace, M. & Saïghi, S. Global parameter estimation of an Hodgkin-Huxley formalism using membrane voltage recordings: application to neuro-mimetic analog integrated circuits. *Neurocomp.* **81,** 75 (2012).
15. Hendrickson, E. B., Edgerton, J. R. & Jaeger, D. The use of automated parameter search to improve ion channel kinetics for neural modeling. *J. Comput. NeuroSci.* **31,** 329 (2011).
16. Reid, M. S., Brown, E. A. & DeWeerth, S. P. A parameterspace search algorithm tested on a Hodgkin-Huxley model. *Biol. Cybern.* **96,** 625 (2007).
17. Pospischil, M. *et al.* Minimal Hodgkin-Huxley type models for different classes of cortical and thalamic neurons. *Biol. Cybern.* **99,** 427 (2008).
18. Vanier, M. C. & Bower, J. M. A comparative study of automated parameter-search methods for compartmental neural models. *J. Comput. Neurosci.* **7,** 149 (1999).
19. Jolivet, R. *et al.* The quantitative single-neuron modeling competition. *Biol. Cybern.* **99,** 417 (2008).
20. Van Geit, W., De Schutter, E. & Achard, P. Automated neuron model optimization techniques: a Review. *Biol. Cybern.* **99,** 241 (2008).
21. Takens, F. Detecting strange attractors in turbulence. *Lect. Notes Math.* **898,** 366 (1981).
22. Nijmeijer, H. & Marels, I. M. Y. An observer look at a synchronization. *IEEE Trans. Circ. Syst. I* **44,** 882 (1997).
23. Vavoulis, D. V., Straub, V. A., Aston, J. A. D. & Feng, J. A self-organizing state-space-model approach for parameter estimation in Hodgkin-Huxley-type models of single neurons. *PLoS Comput. Biol.* **8,** e1002401 (2012).
24. Kitagawa, G. A self-organizing state space model. *J. Am. Stat. Assoc.* **93,** 1203 (1998).
25. Evensen, G. *Data assimilation: the ensemble Kalman filter* (Springer, Berlin 2nd Ed., 2009).
26. Meliza, C. D. *et al.* Estimating parameters and predicting membrane voltages with conductance-based neuron models. *Biol. Cybern.* **108,** 495 (2014).
27. Wächter, A. & Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program. Ser. A* **106,** 25 (2006).
28. Gill, P. E., Murray, W. & Saunders, M. A. SNOPT: a SQP algorithm for large-scale constrained optimization. *SIAM Rev.* **47,** 99 (2005).
29. Byrd, R. H., Gilbert, J. C. & Nocedal, J. A trust region method based on interior point techniques for linear programming. *Math Program.* **89,** 149 (2000).

30. Wächter, A. & Biegler, L. T. Line search filter methods for nonlinear programming: motivation and global convergence. *SIAM J. Opt.* **16,** 1 (2005).
31. Gutenkunst, R. N. *et al.* Universally sloppy parameter sensitivities in systems biology models. *PLoS Comp. Biol.* **1871,** e189 (2007).
32. Marder, E. & Taylor, A. L. Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience* **14,** 133 (2011).
33. Daou, A., Ross, M. T., Johnson, F., Hyson, R. L. & Bertram, R. Electrophysiological characterization and computational models of HVC neurons in Zebra Finch. *J. Neurophysiol.* **110,** 1227 (2013).
34. Abarbanel, H. D. I., Creveling, D. R., Farsian, R. & Kostuk, M. Dynamical state and parameter estimation. *SIAM J. Appl. Dyn. Syst.* **8,** 1341 (2009).
35. Toth, B. A., Kostuk, M., Meliza, C. D., Margoliash, D. & Abarbanel, H. D. I. Dynamical estimation of neuron and network properties I: Variational methods. *Biol. Cybern.* **105,** 217 (2011).
36. Huguenard, J. R. & Mc Cormick, D. A. Simulation of the currents involved in rhythmic oscillations in thalamic neurons. *J. Neurophysiol.* **68,** 1373 (1992).
37. McCormick, D. A. & Huguenard, J. R. A model of the electrophysiological properties of thalamocortical relay neurons. *J. Neurophysiol.* **68,** 1384 (1992).
38. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (CUP, 1994).
39. Tateno, T., Harsch, A. & Robinson, H. P. C. Threshold firing frequency-current relationships of neurons in rat somatosen- sory cortex: Type 1 and Type 2 dynamics. *J. Neurophysiol.* **92,** 2283 (2004).
40. Long, M. A., Jin, D. Z. & Fee, M. S. Support for a synaptic chain model of neuronal sequence generation. *Nature* **468,** 394 (2010).
41. Flath, H. P. *et al.* Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations. *SIAM Sci. Comp.* **33,** 407 (2011).
42. Yuen, K.-V. *Bayesian methods for structural dynamics and civil engineering* (John Wiley and Sons, 2010).
43. O'Leary, T., Sutton, A. C. & Marder, E. Computational models in the age of large datasets. *Curr. Opinion in Neurobiol.* **32,** 87 (2015).
44. Ye, J. X. *et al.* Systematic variational method for statistical nonlinear state and parameter estimation. *Phys. Rev. E* **92,** 052901 (2015).
45. Lovell, P., Carleton, J. B. & Mello, C. V. Genomic analysis of potassium channels genes in songbirds reveal molecular specializations of brain circuits for the maintenance and production of learned vocalizations. *BMC Genomics* **14,** 470 (2013).
46. Warren, W. C. *et al.* The genome of a songbird. *Nature* **464,** 757 (2010).
47. Lovell, P. V., Clayton, D. F., Replogle, K. L. & Mello, C. V. Birdsong Transcriptomics: Neurochemical specializations of the oscine song system. *PLoS One* **3,** e3440 (2008).
48. Wild, J. M., Williams, M. N., Howie, G. J. & Mooney, R. Calcium-binding proteins define interneurons in HVC of the Zebra Finch (taeniopugia guttata). *J. Comparative Neurol.* **483,** 76 (2005).
49. Nixdorf, B. E. Ultrastructural analysis of the development and maturation of synapses and subsynaptic structures in the ectostriatum of the Zebra Finch. *J. Comparative Neurol.* **290,** 472 (1989).
50. Fortune, E. S. & Margoliash, D. Parallel pathways and convergence onto HVC and adjacent neostriatum of adult Zebra Finches (taeniopygia-guttata). *J. Comparative Neurol.* **360,** 413 (1995).
51. Dutar, P., Vu, H. M. & Perkel, D. J. Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVC of the adult Zebra Finch. *J. Neurophysiol.* **80,** 1828 (1998).
52. Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the Zebra Finch. *J. Neuroscience* **20,** 5420 (2000).
53. Lorenz, E. N. Deterministic non periodic flow. *J. Atm. Sci.* **20,** 130 (1963).
54. Gerstner, W. & Kistler, W. M. *Spiking Neuron Models* (CUP, 2002).
55. McCormick, D. A. & Pape, H. C. Properties of a hyperpolarization-activated cation current and its role in rhythmic oscillation in thalamic relay neurons. *J. Physiol. Lond.* **431,** 291 (1990).

Acknowledgements

Author Contributions

A.N. performed calculations, analyzed data and wrote the manuscript. C.D.M. performed electrophysiological experiments and analyzed data. D.M. and H.D.I.A. conceived the experimental procedure and the principles of data assimilation. All authors discussed the findings and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nogaret, A. *et al.* Automatic Construction of Predictive Neuron Models through Large Scale Assimilation of Electrophysiological Data. *Sci. Rep.* **6,** 32749; doi: 10.1038/srep32749 (2016).