Parameter Sensitivity Analysis in Electrophysiological Models Using Multivariable Regression

Eric A. Sobie*
Department of Pharmacology and Systems Therapeutics, Mount Sinai School of Medicine, New York, New York 10029

ABSTRACT Computational models of electrical activity and calcium signaling in cardiac myocytes are important tools for understanding physiology. The sensitivity of these models to changes in parameters is often not well-understood, however, because parameter evaluation can be a time-consuming, tedious process. I demonstrate here what I believe is a novel method for rapidly determining how changes in parameters affect outputs. In three models of the ventricular action potential, parameters were randomized, repeated simulations were run, important outputs were calculated, and multivariable regression was performed on the collected results. Random parameters included both maximal rates of ion transport and gating variable characteristics. The procedure generated simplified, empirical models that predicted outputs resulting from new sets of input parameters. The linear regression models were quite accurate, despite nonlinearities in the mechanistic models. Moreover, the regression coefficients, which represent parameter sensitivities, were robust, even when parameters were varied over a wide range. Most importantly, a side-by-side comparison of two similar models identified fundamental differences in model behavior, and revealed model predictions that were both consistent with, and inconsistent with, experimental data. This new method therefore shows promise as a tool for the characterization and assessment of computational models. The general strategy may also suggest methods for integrating traditional quantitative models with large-scale data sets obtained using high-throughput technologies.

INTRODUCTION

Experimental techniques developed in the past several years have begun to generate enormous quantities of data in large-scale, high-throughput assays. These techniques can assess biologically important quantities such as mRNA abundance, protein-protein interactions, and the activation of transcription factors under a wide range of conditions (1). In many studies, the data generated using these large-scale methods were used to construct ''top-down'' models of interactions between components. Sophisticated techniques derived from graph theory were then applied to analyze the topological characteristics of these models (2,3).

Such models stand in stark contrast, however, to mathematical representations of electrophysiology and ion transport in excitable cells (4,5). These more traditional models are generally constructed using a ''bottom-up'' approach, based on data obtained in quantitative experiments. Unlike results gathered using high-throughput technologies, in which hundreds of factors can change between experimental conditions, the data used to develop mechanistic models are usually obtained by altering experimental variables one at a time.

A wide gap therefore exists between ''high-level'' models that include hundreds or thousands of components but little mechanistic detail, and ''low-level'' models that describe the workings of a handful of elements rather comprehensively. At present, the best methods for relating models based on these contrasting approaches are not clear. Potential strategies for bridging this wide gap are implicit in empirical models that use statistical techniques to relate inputs to outputs (6,7). One such technique, partial least squares (PLS) regression (8), has proven useful in extracting correlations between input variables, possibly noisy and partially redundant, and the outputs of interest. For instance, PLS regression was used to assess the relative importance of diverse cellular signals in predicting apoptosis (9). Such methods have seldom been applied, however, in studies examining the altered expression or function of ion channels, pumps, and transporters. In fact, it is unclear whether a simplified empirical model can provide insights into a cellular process, such as the action potential (AP), that is highly nonlinear and already understood in considerable quantitative detail.

This work began with the premise that simple models generated with PLS regression could be used to relate changes in the expression of ion channels to alterations in physiological phenomena such as APs and cellular calcium transients. As an initial proof of principle, simulated data were generated by randomly varying parameters in computational models of the ventricular myocyte. Somewhat surprisingly, the results indicated that the procedure used, parameter randomization followed by multivariable regression, is a simple yet powerful technique for assessing parameter sensitivity in complex computational models. This method can be used to compare the relative effects of different parameters in determining model output, can identify seemingly counterintuitive behavior, and can highlight differences between models that otherwise appear similar. More broadly, the results suggest that statistical techniques may provide a framework for relating changes in the abundance or function of biophysically important proteins to physiological measures.

Submitted May 6, 2008, and accepted for publication October 28, 2008.

*Correspondence: eric.sobie@mssm.edu

Editor: Arthur Sherman.

0006-3495/09/02/1264/11 $2.00

doi: 10.1016/j.bpj.2008.10.056

METHODS

The goal of the computational modeling was to generate sets of simulated data analogous to those obtained in large-scale gene expression screens, and to relate these to physiological outputs. To produce these simulated data, computations were performed using three models of the mammalian ventricular AP: the ''phase 1'' model of Luo and Rudy (10), and models described more recently by Fox et al. (11) and Kurata et al. (12).

As an initial test, maximum values of ionic conductances in AP models were varied randomly. This procedure was based on the idea that, to a first approximation, an increase in expression of an ion channel will result in a proportional increase in the magnitude of the corresponding ionic current. Two studies identified conditions under which this assumption holds (13,14), although this will likely not be true for all channels and all cell types. Sets of randomly varying conductances were generated by scaling the baseline conductance values listed in the published models. This article adopts the convention that a capital G specifies the maximal conductance of an ionic current, e.g., the sodium current in the phase 1 Luo-Rudy (LR1) model is described as $I_{Na} = G_{Na}m^3hj(V - E_{Na})$. In simulations using newer models (11,12), maximal fluxes through ion pumps and transporters were also scaled. Because the parameters that scale these fluxes are not conductances, these are denoted with Ks. For instance, an increase in Na^+-Ca^{2+} exchange current is achieved by increasing the factor K_{NCX}.

In some simulations (Fig. 5 and the Supporting Material, Fig. S5, Fig. S6, Fig. S7, and Fig. S8), parameters controlling the ionic current gating variables were also varied randomly. For instance, the six ionic currents in the LR1 model are controlled by nine gating variables that range between zero and one. Three of these (K_p, x, n_i) assume new values instantaneously when the membrane potential changes. The temporal evolution of the other six (m, h, j, d, f, n) depends on ''infinity values'' and time constants that both depend on the membrane potential. Auxiliary random variables that either shifted the gating variable infinity curves along the voltage axis, or scaled gating variable time constants at all membrane potentials, were introduced into the LR1 model. The former auxiliary variables are denoted as V (e.g., V_h shifts the infinity curve of the Na^+ current fast inactivation gate), and the latter as p (e.g., a value of $p_d > 1$ slows activation of the slow inward current). Introducing these variables increased the number of randomly varying parameters to 21, 40, and 33 in the models of LR1 (10), Fox et al. (11), and Kurata et al. (12), respectively. Complete lists of the parameters varied and their control values are provided in Table S1, Table S2, and Table S3.

Random scale factors for conductances and time constants were chosen from a log-normal distribution with a median value of 1. Hence the probability that a conductance was doubled equaled the probability that its value was halved. Using this distribution also ensured that the nonphysiological conditions of a negative conductance or negative time constant were never encountered. The parameter σ, specifying the standard deviation of the distribution of log-transformed variables, controlled the extent to which parameters varied. The procedure for generating randomized scale factors is illustrated, along with sample conductance distributions, in Fig. S1. Shifts of the gating variable infinity curves along the voltage axis could be positive or negative, and thus these random variables were normally rather than log-normally distributed.

For each set of randomly chosen conductances, the resulting APs were computed. In simulations with the LR1 model, a 60-s period of quiescence followed by a suprathreshold current pulse was simulated. In simulations with the more recent models, a train of 100 stimuli delivered at 1 Hz was simulated, and the response to the last stimulus in the series was saved for regression analysis. In LR1 simulations, the resting membrane potential (V_{rest}), peak voltage during the AP (V_{peak}), and action potential duration (APD) were recorded for each random set of model parameters. The APD was defined as the time from the maximum upstroke velocity (dV/dt_{max}) to the −60-mV crossing. In simulations with the models of Fox et al. (11) and Kurata et al. (12), the Ca^{2+} transient amplitude ($\Delta[Ca^{2+}]_i$), defined as peak intracellular [Ca^{2+}] minus the level immediately before the AP, was also computed. Simulations were performed with many sets of random parameters to generate input and output matrices. Each input matrix $\mathbf{X}$ had dimensions $n \times p$, where n is the number of sets of random parameters, and p is the number of model parameters varied. Each output matrix $\mathbf{Y}$ had dimensions $n \times m$, where m is the number of outputs.

The PLS regression, performed on matrices $\mathbf{X}$ and $\mathbf{Y}$ using the NIPALS algorithm (8,15), produced a $p \times m$ matrix of regression coefficients $\mathbf{B_{PLS}}$. Given a new set of inputs, this matrix can be used to predict the resulting outputs. In other words, the regression determines $\mathbf{B_{PLS}}$ such that $\mathbf{Y_{predicted}} = \mathbf{X}\times\mathbf{B_{PLS}}$ is close to the original output matrix $\mathbf{Y}$. In the results presented here, parameters were varied independently, and numerous trials were run to obtain an accurate estimate of $\mathbf{B_{PLS}}$. All matrices used for regression therefore had full rank. Because of this, the results obtained using the PLS regression were identical to those that would have been obtained using more traditional multivariable linear regression. For instance, $\mathbf{B}$ can be computed from the formula $\mathbf{B} = (\mathbf{X^T}\times\mathbf{X})^{-1}\times\mathbf{X^T}\times\mathbf{Y}$, where the superscript T denotes the matrix transpose, and the superscript −1 indicates the inverse of a square matrix. PLS regression, rather than standard multivariable regression, was used because this technique can be applied even when the number of variables is much greater than the number of samples (i.e., $p > n$; see Fig. S11). In such a situation, the standard regression formula could not be used because the matrix $\mathbf{X^T}\times\mathbf{X}$ would be singular. However, when many simulations are performed so that $n > p$, the regression coefficients obtained using standard regression and PLS regression are identical.

Before performing regression, values in input and output matrices were mean-centered and normalized by standard deviations, i.e., $x_{new} = (x_{orig} - \mu_x)/\sigma_x$, where μ and σ are computed for each column. Because random-scale factors obeyed a log-normal distribution, most values in the input matrices were log-transformed before computing the values of μ and σ. When normally distributed voltage shifts were included, these values were not log-transformed. Distributions of APD and $\Delta[Ca^{2+}]_i$ were right-skewed, so these values were also log-transformed, and retransformed back into the original units for the display of results. By visual inspection, V_{rest} and V_{peak} were approximately normally distributed; therefore, these variables were not transformed before scaling. Sample output distributions are shown in Fig. S2, and a more thorough description of the input and output matrices, including the normalization and log-transform steps, is provided in Fig. S3.

PLS regression was performed using routines written by Dr. Herve Abdi of the University of Texas at Dallas and available for download at his website (http://wwwpub.utdallas.edu/~herve/). All computations were performed in MATLAB (The MathWorks, Natick, MA).

RESULTS

Fig. 1 *A* shows sample action potentials produced by randomly varying the six maximum ionic conductances in the LR1 model. Normalized values of G_{Na}, G_{si}, G_{K1}, G_K, G_{Kp}, and G_B are given in the bar graph, color-coded to the corresponding AP traces. These random changes could cause APD, V_{rest}, and V_{peak} to increase or decrease, depending on the combination of conductances in a particular trial. This procedure was repeated numerous times to generate input and output matrices, each containing $n = 500$ rows, with $p = 6$ and $m = 3$ columns corresponding to the six conductances and three outputs, respectively. PLS regression produced a $p \times m$ (here, 6×3) matrix of regression coefficients ($\mathbf{B_{PLS}}$). This procedure generates a simple empirical model that can predict the outputs resulting from a new set of input parameters through the formula $\mathbf{Y_{predicted}} = \mathbf{X}\times\mathbf{B_{PLS}}$ (Fig. 1 *B*). Scatterplots (Fig. 2) demonstrate the accuracy of the predictions. For each of the three outputs, the plot shows the ''actual'' value, calculated by the numerical integration

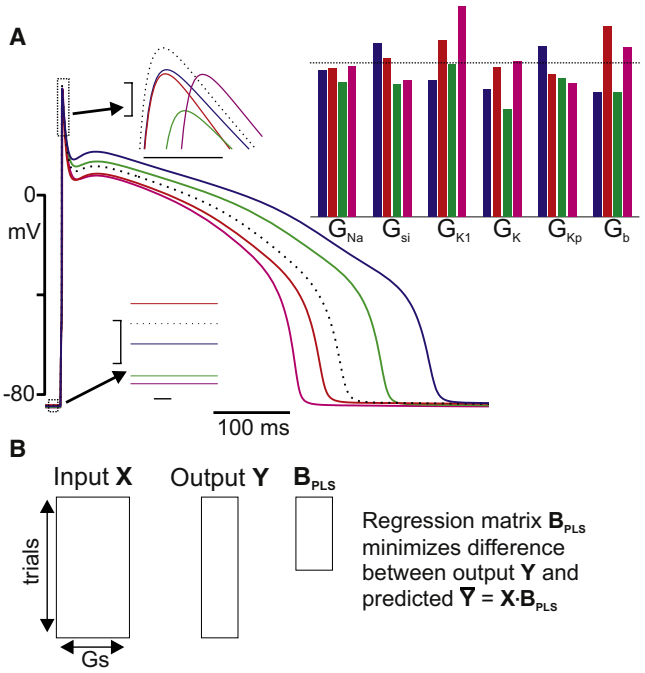

FIGURE 1 (*A*) Changes in AP morphology, compared with control AP (*black dashed line*), produced by randomly varying ionic conductances in the LR1 model. (*Right inset*) Color-coded bar graph shows conductance values, normalized to control levels (*horizontal black line*), of the four trials. (*Top and bottom insets*) Expanded views of AP peaks and resting potentials, respectively. Temporal scale bars, 1 ms; voltage scale bars from 40–42 mV (*top*) and from −85 to −84.5 mV (*bottom*). (*B*) Schematic of input, output, and PLS regression matrix structures.

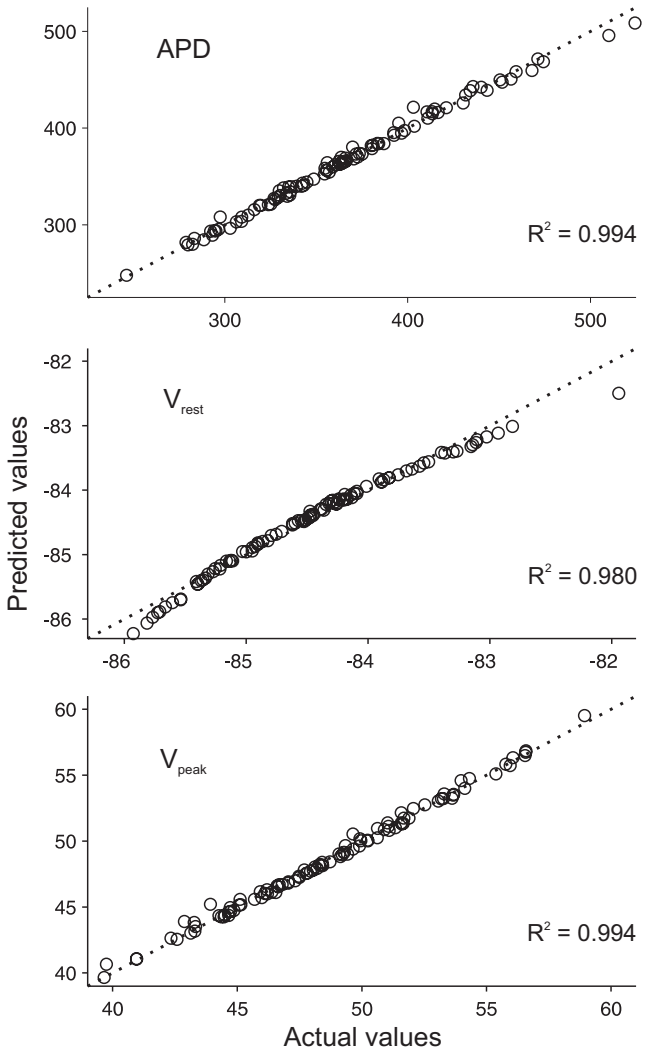

FIGURE 2 Predictions of linear empirical model generated by PLS regression. Scatterplots are displayed for three outputs: APD (*top*), V_{rest} (*middle*), and V_{peak} (*bottom*). Each plot shows the value computed by numerical integration of LR1 equations (*abscissa*) versus the estimate generated by the PLS regression model (*ordinate*). For clarity, only 100 points are shown on each plot, although the regression was performed on a simulated data set containing 500 samples.

of the LR1 differential equations, versus the prediction of the empirical regression model. The large R^2 values (0.98–0.994) indicate that the simplified linear model is highly predictive, despite the nonlinear differential equations in the AP model.

The regression coefficients in the matrix $\mathbf{B_{PLS}}$ indicate how changes in input conductances lead to changes in outputs, with each column reflecting the effects on a particular output. Examining these coefficients allows for an assessment of the relative contributions of the various parameters. For instance, the bar graphs in Fig. 3 *A* indicate that: 1) increases in G_{si}, and decreases in G_K or G_B, lengthen the APD, whereas changes in G_{Na}, G_{K1}, and G_{Kp} have little effect; 2) V_{rest} is determined by a balance between G_{K1} and G_B, with increases in the former causing hyperpolarization, and increases in the latter causing depolarization; and 3) G_{Na} is the only conductance exerting a substantial effect on V_{peak}. This exercise illustrates the potential utility of multivariable regression for parameter sensitivity analyses. Although the effects of changing ionic conductances can easily be determined by varying each parameter individually and recording the results, this procedure becomes time-consuming and tedious for detailed models that consider numerous ion-transport mechanisms. The method described here allows for a simple and rapid evaluation of how changes in parameters affect model outputs. Equally important, this procedure provides a compact graphical representation of the relative effects of each parameter.

The values in $\mathbf{B_{PLS}}$ displayed in Fig. 3 *A* can be understood quantitatively as follows. Because input and output matrices are mean-centered and normalized to standard deviations computed column-by-column, each element of $\mathbf{B_{PLS}}$ is defined relative to the relevant deviations. Thus, a value of 0.5 means that an input value one standard deviation greater than the mean will increase the output by half a standard deviation. Moreover, in the case of APD, both inputs and outputs are log-transformed before computing means and deviations (see Methods). Therefore, both the increase (or decrease) in conductance and the increase (or decrease) in APD can be understood as percent changes relative to the baseline level.

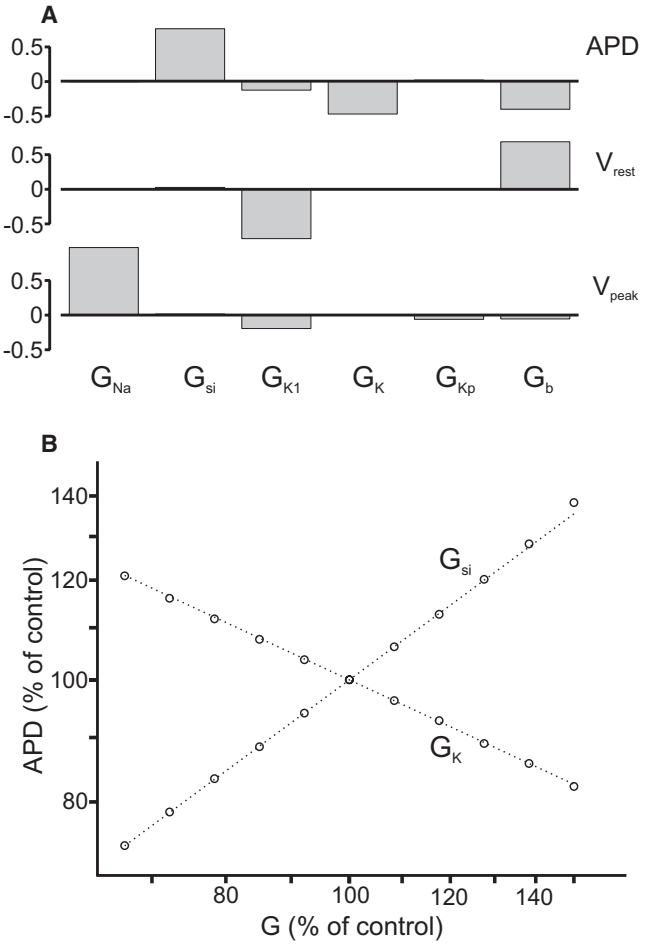

FIGURE 3 (*A*) Columns of matrix B_{PLS} produced by regression algorithm. Elements of B_{PLS} indicate how changes in conductances affect APD (*top*), V_{rest} (*middle*), and V_{peak} (*bottom*). (*B*) Lengthening or shortening of APD caused by changes in G_{si} or G_K, from simulations in which only one parameter was varied. Both simulation results (*symbols*) and predictions of the linear model produced by PLS regression (*dashed lines*) are shown. Axes are plotted on a logarithmic scale.

For instance, the regression coefficient corresponding to the effect of G_{si} on APD is 0.78. This implies that increases in G_{si} of 16% and 34% (one and two standard deviations) increase the APD by 12% and 26% (0.78 and 1.56 standard deviations), respectively. Fig. 3 *B* shows, on a log-log scale and with each variable normalized to control values, plots of APD versus G for simulations in which only G_{si} or G_K was altered. The results of the simulations (*circles*) and the regression model predictions (*dashed lines*) are nearly identical, further demonstrating the accuracy of the empirical model.

In the simulations shown in Figs. 1–3, ionic conductances were varied over a relatively narrow range. The parameter σ, which controls random variation of maximum conductances, was set at 0.15. Hence 95% (± 2 standard deviations) of the randomly chosen conductance values were either increased by $<35\%$ $e^{2(0.15)}$ or decreased by $<26\%$ $e^{-2(0.15)}$. In biological samples, however, the expression levels of ion channels can sometimes change more dramatically. We therefore examined how increasing σ affected the regression model. In these simulations, APD was the only output variable, and values of σ from 0.1–0.5 were tested. Larger values resulted in grossly unphysiological behavior, such as APs that never repolarized, in a significant proportion of trials (>20%; Fig. S3) Fig. 4 *A* shows that the predictive power of the regression model declined somewhat with an increase in σ, with R^2 decreasing from 0.99 to 0.87. However, five of the six regression coefficients in the matrix $\mathbf{B_{PLS}}$ remained essentially constant (Fig. 4 *B*). This demonstrates that when parameters are allowed to vary over a wider range, the parameter sensitivity analysis generated using this procedure is remarkably robust, despite a slight decrease in accuracy.

The single value in $\mathbf{B_{PLS}}$ that did not remain constant with an increase in σ was the regression coefficient corresponding to background conductance G_B. This value increased in magnitude from ~ -0.4 to -0.5 as σ was increased, implying that large changes in G_B have a proportionally greater effect on APD than small changes. Fig. 4 *C*, depicting action potentials simulated with progressively larger values of G_B, shows that this prediction is correct.

In the above simulations, ionic conductances were varied under the assumption that a change in expression of an ion-channel subunit will lead to a change in the magnitude of the ionic current (13,14). However, changes in expression of auxiliary subunits, or posttranslational modifications of channels, can lead to more subtle changes such as shifts in the voltage dependence of gating. To explore whether the randomization/regression procedure could offer insights into these types of effects, additional simulations were performed in which gating characteristics were also varied randomly (see Methods). When random variables that scaled gating variable time constants or shifted gating variable infinity curves were included, the empirical model derived using multivariable regression still generated accurate predictions ($R^2 = 0.97$–0.99; Fig. S5 *A*). Fig. 5 displays the parameter sensitivities that allow for a direct and quantitative comparison between how changes in ionic conductances versus changes in channel gating affect important model outputs. For instance, the values displayed suggest that an 86% increase in inward rectifier conductance G_{K1}, a 33% increase in the delayed rectifier gating variable (n) time constant, and a 2.75-mV shift of the slow inward current activation variable (d) will all cause a 30-ms decrease in APD. These predictions are correct to within 5 ms (Fig. S5, *B*), illustrating the quantitative accuracy of the regression model.

A question of interest was whether, compared with scaling conductances, the accuracy of the linear regression became compromised more easily when gating parameters were randomly varied. To address this question, parameter variability, controlled by σ, was systematically increased as in Fig. 4. Three sets of simulations were performed, using only Gs, only ps, or only Vs as the randomly varying parameters (Fig. S6). Somewhat surprisingly, the decline in accuracy of

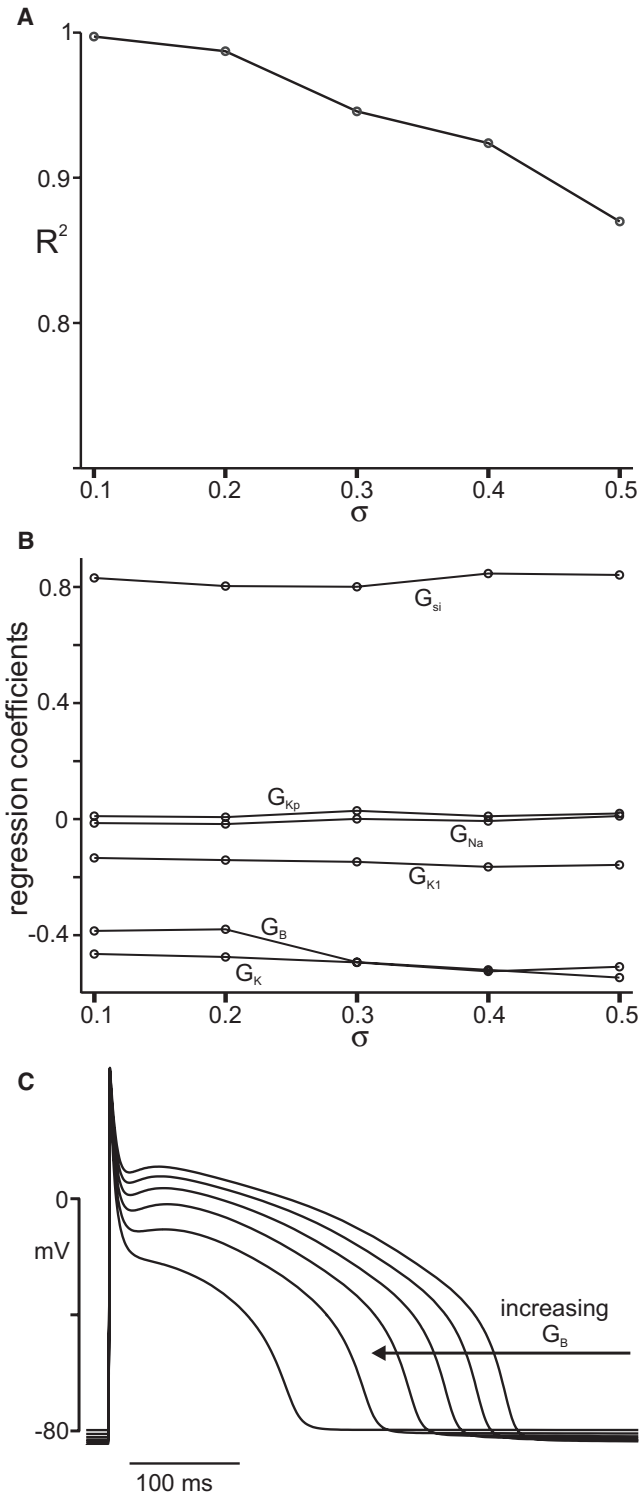

FIGURE 4 Effects of changes in ionic conductance variation. Simulations were performed with the LR1 model at different values of σ, the parameter controlling the spread of random conductance distributions. (*A*) The ability of the PLS model to predict APD, quantified by R^2, decreased but remained fairly high ($\geq$0.87), as σ increased from 0.1 to 0.5. (*B*) Coefficients in the regression matrix B_{PLS} remain essentially constant at all values of σ, with the exception of the value corresponding to G_B, which is smaller in magnitude at $\sigma < 0.3$ than at greater values. (*C*) Simulated APs at progressively larger values of G_B, with all other maximal conductances kept constant at their control values. The increased relative shortening seen at larger values confirms the prediction of the PLS model. The APs shown correspond to logarithmically spaced values of G_B: 100%, 120%, 144%, 173%, 207%, and 249% of the control level.

the empirical regression model with an increase in σ was roughly equivalent in all three cases. Significantly, the derived regression coefficients remained essentially constant at all values of σ whether conductances, gating variable time constants, or infinity curves were randomly varied. These results demonstrate that a model's sensitivity to ''nonlinear'' parameters can also be assessed using this regression procedure.

The feasibility of the method was initially tested using the LR1 model, because the few conductances in this formulation have well-understood effects on the AP. To verify that this procedure can provide new physiological insights, simulations were also performed with the ventricular myocyte models described by Fox et al. (11) and Kurata et al. (12). Conductances were randomized, and regression was performed on the results as before, in this case using APD and Ca^{2+} transient amplitude ($\Delta[Ca^{2+}]_i$) as outputs. The APs produced by these models with control parameters, and bar graphs showing the regression coefficients, are given in Fig. 6. In a separate set of simulations, both conductances and gating variable parameters were varied, analogous to the LR1 simulations in Fig. 4. Although the discussion below focuses on the effects of conductance changes, the full set of results is provided for completeness in Fig. S7 and Fig. S8.

Because the models of Fox et al. (11) and Kurata et al. (12) contain many of the same ionic currents and produce APs with similar morphologies, it is instructive to compare the regression coefficients. Comparing the bar graphs for APD in the left and right columns of Fig. 6 reveals some surprises. The plots suggest that: 1) changes in L-type Ca^{2+} current have a much larger relative effect in the model of Kurata et al. (12) than in the model of Fox et al. (11), where the effects are similar to those produced by changes in background Na^+ current; 2) changes in G_{K1} have a much larger effect than changes in G_{Kr} or G_{Ks} in the model of Fox et al. (11), whereas the effects of these three currents are similar in the model of Kurata et al. (12). Fig. 7, *A* and *B*, shows the results of additional simulations that confirm these predictions. With respect to $\Delta[Ca^{2+}]_i$, the regression analysis suggests that increasing the maximum rate of sarcoplasmic reticulum Ca^{2+} release (K_{rel}) causes a decrease in steady-state $\Delta[Ca^{2+}]_i$ in the model of Fox et al. (11), but an increase in the model of Kurata et al. (12). This unexpected finding, in which changing a single model parameter causes opposite effects in two similar models, was confirmed in additional simulations (Fig. 7 *C*). These results show that the new parameter sensitivity analysis method can be used to identify nonobvious but potentially important differences between models. The differences can then be compared with experimental data to identify aspects of particular models that need to be modified, or, when existing data are unavailable,

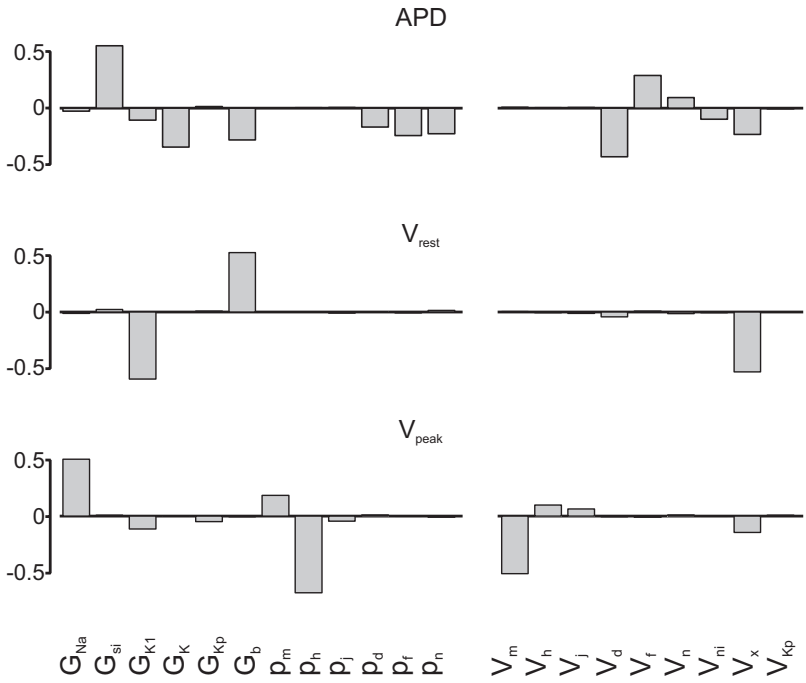

FIGURE 5 PLS regression results with ionic current gating variables included in the analysis. As described in the text, auxiliary variables that either scaled the gating variable time constants (p) or shifted the gating variable infinity curves (V) were introduced into the LR1 model. Bar graphs display regression coefficients. Each value indicates how changing a given model parameter affects APD (*top*), V_{rest} (*middle*), or V_{peak} (*bottom*). Abbreviations are as follows: p_m, p_h, p_j, p_d, p_f, and p_n, respectively, scale the following time constants: Na^+ current activation, Na^+ current fast inactivation, Na^+ current slow inactivation, slow inward current activation, slow inward current inactivation, and delayed rectifier K^+ current activation. The variables V_m, V_h, V_j, V_d, V_f, and V_n shift these gating variable infinity curves along the voltage axis. In addition, the variables V_{ni}, V_x, and V_{Kp} shift along the voltage axis instantaneous gating variables that affect delayed rectifier K^+ current, inward rectifier K^+ current, and plateau K^+ current, respectively.

the computational results can suggest new experiments to clarify discrepancies (see Discussion).

Finally, we sought to determine whether this new method could be used to analyze a model output that is both nonlinear and clinically relevant, i.e., the susceptibility of the model cell to APD alternans. Similar to most cardiac AP models, as well as real cardiac cells, the model of Fox et al. (11) shows a constant APD from beat to beat at slow pacing rates, but a short-long alternating pattern, called APD alternans, at faster rates (Fig. 8 *A*). In this case, defining a single APD is inappropriate, because it changes from one stimulus to the next. However, the initial pacing rate at which alternans is induced, which depends on the model parameters, can be collected as an output and analyzed using regression. The regression coefficients in Fig. 8 *B* illustrate how changes in ionic conductances affect the basic cycle length at which alternans first occurs (BCL_{alt}). A positive value means that increasing a particular conductance increases BCL_{alt}, i.e., this change makes the cell more susceptible to alternans. These parameter sensitivities are consistent with the results of Fox et al. (11) in several ways. For instance, Fox et al. (11) showed that when L-type Ca^{2+} current magnitude was reduced, the cell became less susceptible to alternans. They also documented a resistance to APD alternans through a relatively small increase in G_{K1} (7%), a larger increase in G_{Kr} (62%), or a much larger increase in G_{Ks} (14 times). The relative magnitudes of the parameter sensitivities in Fig. 8 *B* are consistent with these previous findings. More importantly, the analysis identified two additional model parameters not previously examined by Fox et al. (11) that are predicted to exert large effects on a cell's vulnerability to alternans. These results in Fig. 8, *C* and *D*, confirm the predictions of the regression model. If the background Na^+ conductance G_{Nab} is increased by 25%, the cell becomes more likely to develop alternans, such that they occur at a BCL of 235 ms (Fig. 8, *C*, *left*). A decrease in G_{Nab} by 20% makes the cell resistant to alternans, such that APD is still stable at a BCL of 170 ms (Fig. 8, *C*, *right*). In contrast, changes in K_{NaK}, which controls the maximum turnover rate of the Na^+-K^+ ATPase, have opposite effects. Blocking the pump current makes the cell more susceptible (Fig. 8, *D*, *left*), and increasing K_{NaK} makes the cell less susceptible (Fig. 8, *D*, *right*), to APD alternans. These results demonstrate that empirical linear regression models can identify unexpected model behaviors, even when considering nonlinear model outputs.

DISCUSSION

In this study, parameters in computational models of the ventricular myocyte were randomized, and numerous simulations were run with different combinations of parameters. Computations were performed with both an extremely well-characterized model (10) and two newer models that have been less extensively studied (11,12). Using 1), the randomized parameters, and 2), the AP characteristics computed from the simulations, as input and output matrices, respectively, the results were subjected to multivariable regression. Input parameters included both maximal conductances of ionic currents and factors controlling ion-channel gating, and outputs included physiologically important measures such as APD and Ca^{2+} transient amplitude. The randomization/regression procedure produced an empirical, linear model that quantified correlations between parameters and outputs.

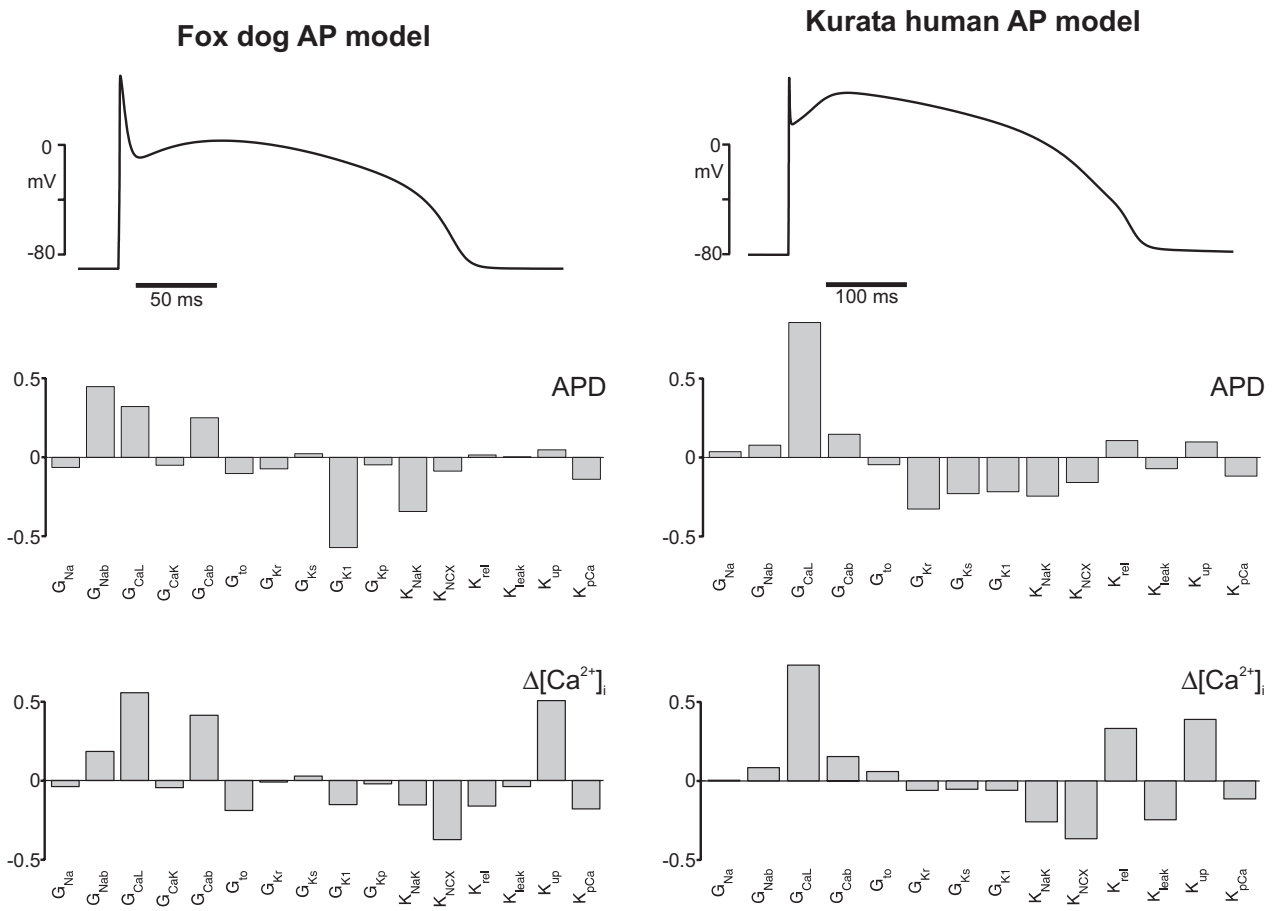

FIGURE 6 Parameter sensitivity analysis of newer AP models. (*Left* and *right*, *top*) Simulated APs with control parameter values in models of Fox et al. (11) and Kurata et al. (12), respectively. Bar graphs show regression coefficients in the matrix B_{PLS}, indicating how changes in model parameters affect APD (*top graphs*) and $\Delta[Ca^{2+}]_i$ (*bottom graphs*). Complete lists of model parameters randomly varied are provided in Table S2 and Table S3.

Despite the many nonlinear equations of the electrophysiological models, the predictive power of the regression models was quite strong. Moreover, the regression coefficients reflect parameter sensitivity, i.e., each coefficient indicates the change in a model output produced by increasing or decreasing a parameter. When these coefficients were examined in two seemingly similar models, the analysis identified differences in model behavior that likely would have remained hidden without extensive testing. The computational results presented therefore suggest that this new method, i.e., randomization of

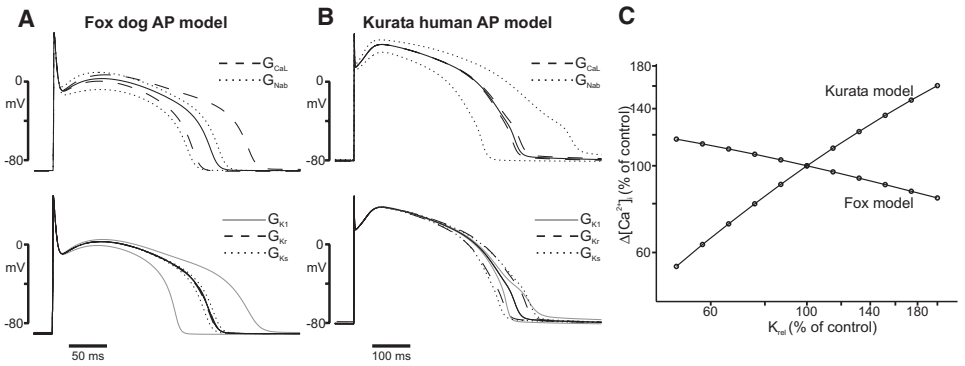

FIGURE 7 Verification of regression model predictions. (*A* and *B*) APs generated with control conductance values (*black line*), and those with a 50% increase and a 33% decrease in a particular conductance. (*A*) Changes in G_{CaL} (*dashed traces*) and G_{Nab} (*dotted traces*) in the model of Fox et al. (11) of dog AP are shown at top. Because these conductances correspond to depolarizing currents, an increase in either lengthens the AP, whereas a decrease causes shortening. (*Bottom plots*) Changes in G_{K1} (*gray*), G_{Kr} (*dashed*), and G_{Ks} (*dotted*). Increases in these repolarizing currents shorten APs, and vice versa. (*B*) Changes in G_{CaL} and G_{Nab} in the model by Kurata et al. (12) of human AP are shown at top. Changes in G_{K1}, G_{Kr}, and G_{Ks} are at bottom. The relative changes observed are all consistent with predictions generated by performing PLS regression on sets of simulated data. (*C*) Confirmation of PLS model predictions regarding simulated Ca^{2+} transients. The plot shows changes in Ca^{2+} transient amplitude ($\Delta[Ca^{2+}]_i$), normalized to control amplitude, produced by changes in K_{rel} in the two models as labeled. Changing this parameter caused opposite effects in the two models, as implied by the comparison of regression coefficients in Fig. 6. Axes are on a logarithmic scale.

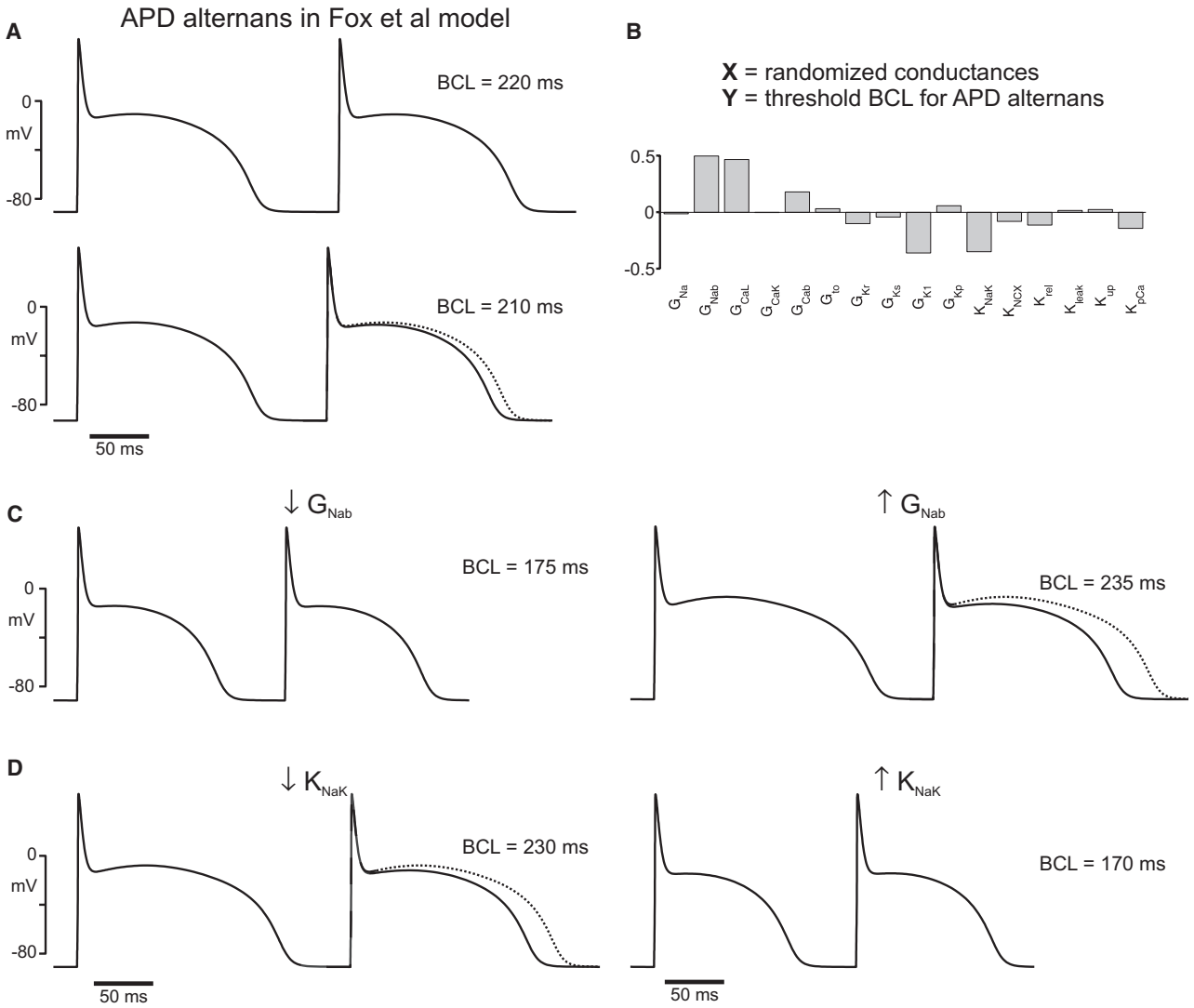

FIGURE 8 Use of PLS regression to predict susceptibility to APD alternans. Simulations were performed with the by Fox et al. (11) of canine action potential (11). The protocol was to pace the myocyte at a relatively long basic cycle length (BCL), and progressively decrease BCL until APD alternans were induced. To avoid transient effects contaminating the results, 100 electrical stimuli were delivered at each BCL, and the 99th and 100th APDs were compared. For each set of randomly chosen ionic conductances, the threshold BCL for alternans (BCL_{alt}) was collected as an output. (*A*) With control parameters, APD is constant at a BCL of 220 ms, but a BCL of 210 ms results in APD alternans. The 99th AP (*dashed line*) is superimposed on the 100th AP (*solid line*) to illustrate the alternation. (*B*) Regression coefficients obtained using ionic conductances as inputs, and BCL_{alt} as output. A positive value means that increasing a conductance makes the model cell more susceptible to alternans. (*C*) Confirmation of regression model prediction regarding background Na^+ conductance G_{Nab}. Consistent with regression coefficients in *B*, a 20% decrease in G_{Nab} makes the cell less susceptible to alternans, such that APD is stable at a BCL of 175 ms (*left*). A 25% increase in G_{Nab} causes alternans to appear at a slower BCL of 235 ms (*right*). (*D*) Confirmation of regression model Na^+-K^+ pump rate (K_{NaK}) prediction. A 20% decrease in K_{NaK} makes the model cell more susceptible to APD alternans (*left*), whereas a 25% increase in K_{NaK} makes the cell resistant to alternans (*right*).

parameters followed by regression, can be useful in the evaluation of cardiac electrophysiological models. This type of analysis, in which the effects of many parameters are determined but mechanistic information is not provided directly, complements the methods used in previous comparisons of cardiac computational models (16,17), where differences in mechanisms were examined closely, but the effects of many parameters were not investigated.

Applying this new procedure to the recent ventricular myocyte models of Fox et al. (11) and Kurata et al. (12) uncovered some surprising and potentially important behaviors. One set of predictions concerned the relative effects on APD of changes in different K^+ currents. Each model displayed some strengths and some weaknesses in terms of its consistency with experimental data. It is well-established in dogs (18–20), and somewhat less well-established in humans (21), that blocking rapid delayed rectifier current (I_{Kr}) causes a greater prolongation of APD than blocking slow delayed rectifier current (I_{Ks}), which has little effect. The predictions of the model of Fox et al. (11) were

consistent with this observation, whereas the model of Kurata et al. (12), which showed nearly equal $\mathbf{B_{PLS}}$ coefficients corresponding to these two conductances, generated inconsistent predictions. On the other hand, the analysis demonstrated that changes in inward rectifier current (I_{K1}) have a much greater effect on APD than changes in I_{Kr} in the model of Fox et al. (11). This prediction is inconsistent with experiments in dog ventricular muscle showing that block of I_{Kr} with dofetilide caused greater APD prolongation than block of I_{K1} with 10 μM Ba^{2+} (18). Interestingly, when the effects of parameter changes on APD were assessed at both slow and fast pacing rates, neither model adequately reproduced the ''reverse rate dependence'' of APD prolongation caused by I_{Kr}-blocking drugs such as dofetilide (Fig. S9 and Fig. S10). In a model exhibiting reverse rate dependence, the magnitude of the G_{Kr} coefficient will be greater at the slow compared with the fast pacing rate. The coefficients are comparable at both rates in either model, however, indicating that neither shows reverse rate dependence. In quantitative terms, the regression analyses predict that a 70% reduction in G_{Kr} lengthens APD by 11% and 13% (slow and fast) in the model of Fox et al. (11), and by 34% and 31% (slow and fast) in the model of Kurata et al. (12).

Another interesting prediction identified by the regression analysis was that, in the two models, changes in the maximal rate of sarcoplasmic reticulum Ca^{2+} release (K_{rel}) led to opposite effects on the steady-state Ca^{2+} transient amplitude ($\Delta[Ca^{2+}]_i$). This result is of interest because the intracellular channel responsible for Ca^{2+} release, the ryanodine receptor, is considered a potential therapeutic target for diseases such as heart failure (22,23). Voltage-clamp studies showed that altering the ryanodine-receptor open probability does not cause sustained changes in $\Delta[Ca^{2+}]_i$ because of ''autoregulation'' of the system (24,25). If the beat-to-beat voltage waveform is not fixed, however, the situation is more complicated. In this case, short-term changes in sarcoplasmic reticulum Ca^{2+} release can affect the AP shape, alter the balance of Ca^{2+} across the cell membrane, and lead to either an increase or a decrease in steady-state Ca^{2+} transient amplitude. Thus, the prediction of either model is possible, and, to my knowledge, experiments to distinguish between the two possibilities have not been performed in canine or human ventricular myocytes. This issue can be resolved by pacing an isolated ventricular myocyte at a constant rate and comparing steady-state $\Delta[Ca^{2+}]_i$ in control conditions to that seen after adding a low dose of caffeine (to sensitize ryanodine receptors) or tetracaine (to inhibit ryanodine receptors).

Regression methods were also used to gain insights into factors that influence the development of APD alternans (Fig. 8). The simulations presented by Fox et al. (11) identified how changes in L-type Ca^{2+} current, I_{Kr}, or I_{Ks} could either increase or decrease a cell's propensity to develop alternans (11). Our analysis confirmed these results and identified two additional model parameters that greatly affect the alternans vulnerability of the model myocyte. One of these in particular, the maximum rate of the Na^+-K^+ ATPase (K_{NaK}), has potential clinical relevance, because the heart-failure drug digoxin inhibits this pump. Early experimental results obtained in a multicellular preparation suggest that digoxin can suppress mechanical alternans (26), a finding at odds with the present computational prediction. To my knowledge, however, this result has not been confirmed with microelectrode measurements of APs, or with more contemporary techniques such as multisite optical mapping. Because several models similarly predict that inhibiting I_{NaK} increases alternans susceptibility (not shown), this appears to be another model prediction that requires additional experiments to confirm or refute.

Given the nonlinearities in the electrophysiology models, it may seem surprising that linear regression models can have strong predictive power. This result may be especially unexpected when inputs such as gating variable time constants, or outputs such as alternans pacing thresholds, are analyzed. To understand this apparent paradox, it is important to consider what assumptions are made, and what assumptions are not made, in deriving the empirical linear models. The regression procedure does not assume, or implicitly calculate, a linear approximation to the AP model's differential equations. Instead, the regression only calculates a linear approximation to an ''input-output'' relationship that is not known a priori. Thus, only empirical tests can determine whether the linear approximation will be valid over a narrow range or wide range of parameter values. The literature demonstrates that simple approximations to input-output relationships may be highly predictive, even if the biological phenomena are complex and nonlinear. Historic examples include Hill's force-velocity relationship in muscle (27) and the strength-duration curve describing the threshold for electrical excitation (28). More recently, linear regression models provided novel insights into the signaling pathways that control apoptosis (9). The regression models presented here are in the tradition of such empirical, ''data-driven'' models. However, these regression models will not necessarily generate accurate predictions under all circumstances. In particular, when pathological conditions increase the likelihood that a myocyte will develop an early afterdepolarization, it is likely that small changes in ionic currents will lead to large changes in APD. In this case, action potentials that show early afterdepolarizations will be expected to deviate substantially from the predictions of such regression models.

The elements of the regression matrix $\mathbf{B_{PLS}}$ in this study are similar to response coefficients calculated using metabolic control analysis (MCA). This technique, used extensively in steady-state analyses (29), was recently extended to analyze how changes in parameters affect outputs such as oscillation frequencies (30,31). Both MCA and this new approach quantify how variations in model parameters lead to changes in outputs. A key difference, however, is that response coefficients in MCA are computed by making very small changes

to parameters, under the assumption that systems behave linearly with small perturbations. In contrast, the approach outlined here is completely empirical and does not require any such implicit assumptions. Parameters can be varied over a narrow range, or over a wider range, depending on which is likely to offer the greatest biological insight. A disadvantage of this purely empirical approach, however, is that the parameter sensitivities have no formal mathematical relationship to the dynamical model's differential equations.

A model's parameter sensitivity can also be assessed by varying each parameter systematically, and performing thousands to millions of simulations to explore the parameter space thoroughly (32). A technique such as ''dimension stacking'' can then be used to visualize the results (33,34). This approach clearly provides more information than the method outlined here, which only considers a neighborhood within the parameter space. The main advantage of the current strategy is the computational cost: parameter sensitivities can be assessed by performing only a few hundred simulations with randomly varying parameters. An additional advantage of the new method is that it provides for a compact and intuitive display of the results. The images generated using dimension stacking (33,34), although rich in information, require experience to interpret, compared with the bar graphs in this study.

The results here suggest future applications beyond the evaluation of additional computational models. One would be for calculated parameter sensitivities to guide the development of new models. Mathematical representations of APs and Ca^{2+} transients in heart cells are rarely built de novo. Instead, these models typically evolve, as investigators modify parameters and equations in an existing model to match the behavior in the cell type or condition of interest (12,35,36). Although this process is constrained by experimental observations, the data are usually incomplete, and the set of parameters that recapitulates the data cannot be determined uniquely. Model development therefore involves tuning parameters iteratively to match a particular target AP or Ca^{2+} transient morphology. Regression analysis can assist this process by identifying the parameter changes that move model output toward the target, and, just as important, by showing which parameters have little effect.

A second application would be to analyze, with PLS regression, correlations between physiological outputs and gene expression. In this case, for instance, each column in $\mathbf{X}$ would correspond to the abundance of an ion-channel transcript obtained from a large-scale expression screen (37,38), and the rows of $\mathbf{X}$ would be formed by performing the screen under different conditions. The outputs in $\mathbf{Y}$ would be experimental data: either cellular characteristics such as APD, or higher-level measurements such as hemodynamic or electrocardiogram variables. An empirical model generated using regression could then predict the outputs that would be associated with a different pattern of gene expression. In this circumstance, the number of variables is likely to be much greater than the number of experimental samples. Although this condition precludes the use of standard regression formulas because of the singularity of the matrix $\mathbf{X^T}\times\mathbf{X}$, PLS regression is well-suited to such ''large p, small n'' problems (8,15). Fig. S11 demonstrates the use of PLS regression in such a situation. These simulations, performed with the model of Fox et al. (11), show that even if the number of variables ($p = 16$) is greater than the number of samples ($n = 12$), PLS still provides reasonable estimates of regression coefficients. For analyses of computational models, of course, excellent estimates of regression coefficients can always be obtained by running additional simulations to generate more samples.

In conclusion, this study describes what I believe is a new method that allows for the rapid evaluation of parameter sensitivities in complex computational models. By randomizing parameters, running multiple trials, and performing regression, a user can quickly determine which model parameters play the largest role in determining specific outputs. Comparing the results between models and to experimental data can identify unexpected or unrealistic behavior in existing models, which can suggest modifications to the models or new experiments to clarify points of disagreement. The utility of this approach was demonstrated using ventricular action potential models, but the strategy may be applied to other categories of computational models. In the longer term, these results may provide insights into how best to integrate detailed quantitative knowledge of physiology with new data obtained using high-throughput techniques.

SUPPORTING MATERIAL

Tables and figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(08)03225-6.

This work was supported by National Institutes of Health grants HL076230 and GM071558. The author thanks W. A. Coetzee, R. Iyengar, W. J. Lederer, J. A. Wasserstrom, K. R. Laurita, and A. W. Trafford for helpful discussions.

REFERENCES

1. Hood, L., J. R. Heath, M. E. Phelps, and B. Lin. 2004. Systems biology and new technologies enable predictive and preventative medicine. *Science*. 306:640–643.
2. Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, et al. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*. 98:4569–4574.
3. Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*. 403:623–627.
4. Rudy, Y., and J. R. Silva. 2006. Computational biology in the study of cardiac ion channels and cell electrophysiology. *Q. Rev. Biophys.* 39:57–116.
5. Noble, D., and Y. Rudy. 2001. Models of cardiac ventricular action potentials: iterative interaction between experiment and simulation. *Philos. Trans. R. Soc. Lond. [A] Math. Phys. Eng. Sci.* 359:1127–1142.
6. Ideker, T., and D. Lauffenburger. 2003. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* 21:255–262.

7. Janes, K. A., and D. A. Lauffenburger. 2006. A biological approach to computational models of proteomic networks. *Curr. Opin. Chem. Biol.* 10:73–80.
8. Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression—a tutorial. *Anal. Chim. Acta.* 185:1–17.
9. Janes, K. A., J. G. Albeck, S. Gaudet, P. K. Sorger, D. A. Lauffenburger, et al. 2005. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science.* 310:1646–1653.
10. Luo, C. H., and Y. Rudy. 1991. A model of the ventricular cardiac action potential. Depolarization, repolarization, and their interaction. *Circ. Res.* 68:1501–1526.
11. Fox, J. J., J. L. McHarg, and R. F. Gilmour, Jr. 2002. Ionic mechanism of electrical alternans. *Am. J. Physiol. Heart Circ. Physiol.* 282:H516–H530.
12. Kurata, Y., I. Hisatome, H. Matsuda, and T. Shibamoto. 2005. Dynamical mechanisms of pacemaker generation in I_{K1}-downregulated human ventricular myocytes: insights from bifurcation analyses of a mathematical model. *Biophys. J.* 89:2865–2887.
13. Schulz, D. J., J. M. Goaillard, and E. Marder. 2006. Variable channel expression in identified single and electrically coupled neurons in different animals. *Nat. Neurosci.* 9:356–362.
14. Schulz, D. J., J. M. Goaillard, and E. E. Marder. 2007. Quantitative expression profiling of identified neurons reveals cell-specific constraints on highly variable levels of gene expression. *Proc. Natl. Acad. Sci. USA.* 104:13187–13191.
15. Abdi, H. 2007. Partial least squares (PLS) regression. *In* Encyclopedia of Measurement and Statistics. N. J. Salkind, editor. Sage, Thousand Oaks, CA, pp. 740–744.
16. Cherry, E. M., and F. H. Fenton. 2007. A tale of two dogs: analyzing two models of canine ventricular electrophysiology. *Am. J. Physiol. Heart Circ. Physiol.* 292:H43–H55.
17. Ten Tusscher, K. H., O. Bernus, R. Hren, and A. V. Panfilov. 2006. Comparison of electrophysiological models for human ventricular cells and tissues. *Prog. Biophys. Mol. Biol.* 90:326–345.
18. Biliczki, P., L. Virag, N. Iost, J. G. Papp, and A. Varro. 2002. Interaction of different potassium channels in cardiac repolarization in dog ventricular preparations: role of repolarization reserve. *Br. J. Pharmacol.* 137:361–368.
19. Varro, A., B. Balati, N. Iost, J. Takacs, L. Virag, et al. 2000. The role of the delayed rectifier component I_{Ks} in dog ventricular muscle and Purkinje fibre repolarization. *J. Physiol.* 523:67–81.
20. Volders, P. G., M. Stengl, J. M. van Opstal, U. Gerlach, R. L. Spatjens, et al. 2003. Probing the contribution of I_{Ks} to canine ventricular repolarization: key role for beta-adrenergic receptor stimulation. *Circulation.* 107:2753–2760.
21. Jost, N., L. Virag, M. Bitay, J. Takacs, C. Lengyel, et al. 2005. Restricting excessive cardiac action potential and QT prolongation: a vital role for I_{Ks} in human ventricular muscle. *Circulation.* 112:1392–1399.
22. Venetucci, L. A., A. W. Trafford, M. E. Diaz, S. C. O'Neill, and D. A. Eisner. 2006. Reducing ryanodine receptor open probability as a means to abolish spontaneous Ca^{2+} release and increase Ca^{2+} transient amplitude in adult ventricular myocytes. *Circ. Res.* 98:1299–1305.
23. Wehrens, X. H., S. E. Lehnart, and A. R. Marks. 2005. Ryanodine receptor-targeted anti-arrhythmic therapy. *Ann. N. Y. Acad. Sci.* 1047:366–375.
24. Eisner, D. A., H. S. Choi, M. E. Diaz, S. C. O'Neill, and A. W. Trafford. 2000. Integrative analysis of calcium cycling in cardiac muscle. *Circ. Res.* 87:1087–1094.
25. Trafford, A. W., M. E. Diaz, G. C. Sibbring, and D. A. Eisner. 2000. Modulation of CICR has no maintained effect on systolic Ca^{2+}: simultaneous measurements of sarcoplasmic reticulum and sarcolemmal Ca^{2+} fluxes in rat ventricular myocytes. *J. Physiol.* 522:259–270.
26. Badeer, H. S., U. Y. Ryo, W. F. Gassner, E. J. Kass, J. Cavaluzzi, et al. 1967. Factors affecting pulsus alternans in the rapidly driven heart and papillary muscle. *Am. J. Physiol.* 213:1095–1101.
27. Hill, A. V. 1938. The heat of shortening and the dynamic constants of muscle. *Proc. R. Soc. Lond. [B] Biol. Sci.* 126:136–195.
28. Geddes, L. A., and J. D. Bourland. 1985. Tissue stimulation—theoretical considerations and practical applications. *Med. Biol. Eng. Comput.* 23:131–137.
29. Fell, D. A. 1992. Metabolic control analysis—a survey of its theoretical and experimental development. *Biochem. J.* 286:313–330.
30. Ingalls, B. P., and H. M. Sauro. 2003. Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *J. Theor. Biol.* 222:23–36.
31. Reijenga, K. A., H. V. Westerhoff, B. N. Kholodenko, and J. L. Snoep. 2002. Control analysis for autonomously oscillating biochemical networks. *Biophys. J.* 82:99–108.
32. Prinz, A. A., C. P. Billimoria, and E. Marder. 2003. Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. *J. Neurophysiol.* 90:3998–4015.
33. Matveev, V., R. Bertram, and A. Sherman. 2006. Residual bound Ca^{2+} can account for the effects of Ca^{2+} buffers on synaptic facilitation. *J. Neurophysiol.* 96:3389–3397.
34. Taylor, A. L., T. J. Hickey, A. A. Prinz, and E. Marder. 2006. Structure and visualization of high-dimensional conductance spaces. *J. Neurophysiol.* 96:891–905.
35. Bernus, O., R. Wilders, C. W. Zemlin, H. Verschelde, and A. V. Panfilov. 2002. A computationally efficient electrophysiological model of human ventricular cells. *Am. J. Physiol. Heart Circ. Physiol.* 282: H2296–H2308.
36. Wang, L. J., and E. A. Sobie. 2008. Mathematical model of the neonatal mouse ventricular action potential. *Am. J. Physiol. Heart Circ. Physiol.* 294:H2565–H2575.
37. Gaborit, N., S. Le Bouter, V. Szuts, A. Varro, D. Escande, et al. 2007. Regional and tissue specific transcript signatures of ion channel genes in the non-diseased human heart. *J. Physiol. (Lond.).* 582:675–693.
38. Harrell, M. D., S. Harbi, J. F. Hoffman, J. Zavadil, and W. A. Coetzee. 2007. Large-scale analysis of ion channel gene expression in the mouse heart during perinatal development. *Physiol. Genom.* 28:273–283.