Multiple Regression Analysis of Parameter Sensitivity in Ventricular Myocyte Model

Yuan Hung Lo
A. Michailova, Cardiac Research Group, UC San Diego, La Jolla, California
D. Abramson, MeSsAGE Lab, Monash University, Victoria, Australia

ABSTRACT: electrophysiological computational models of cardiac ventricular myocyte are often times complex and involve numerous parameters and component interactions. Therefore, a viable method is required to assess model stability and to learn about the physiological interactions between various parameters. A holistic study of cardiac myocyte may explain important disease pathologies and offer insights into potential treatment options. Since an *in silico* experiment is easy to manipulate and allows the investigator to collect large-scale data sets, one can use statistical methods to study model outcomes and qualitatively evaluate parameter sensitivity in relation to the outputs. Furthermore, statistical methods can correlate multiple variables and identify novel interactions that are oftentimes elusive in *in vitro* experiments. Sobie (2009) suggested the use of partial least square regression as an exploratory analysis tool in analysing parameter sensitivity in electrophysiological models. He demonstrated its potential as a viable method for characterizing and assessing computational models. In this report I extend the use of PLS regression – in conjunction with Nimrod/E experimental design – as theory confirmation tools to assess Shannon-Bers ventricular myocyte model and also to examine new features of the model or shortfalls that could compromise its modelling capabilities.

NOTATIONS

G_{CaL}	SL L-type Ca channel conductance
G_{K1}	SL inward rectifier K channel conductance
K_{rel}	SR Ca-induced Ca release channel CICR conductance
K_{up}	SR Ca uptake (SERCA) channel (Ca ATPase) conductance
G_{toslow}	SL Transient outward K channel conductance
G_{Kr}	SL rapid delayed rectifier K current conductance
G_{Clbk}	SL background Cl (leak) channel conductance
K_{NaK}	SL Na-K exchanger current (low sensitivity) conductance
K_{NCX}	SL Na-Ca exchanger current conductance
G_{Na}	Na conductance
G_{Nab}	Na background conductance
G_{CaK}	K permeability of L-type Ca current
G_{CaNa}	Na permeability of L-type Ca conductance
G_{Cab}	Ca background conductance
G_{to}	transient outward K conductance
G_{Ks}	slowed delayed inward rectifier K conductance
G_{Cl}	Ca-activated Cl current conductance
K_{leak}	passive SR Ca leak scaling factor
K_{pCa}	SL Ca pump current conductance

BACKGROUND

Partial Least Square Regression (PLSR)

PLSR is a statistical method similar to principal component regression that combines features from and generalizes principal component analysis (PCA) and multiple linear regression (MLR). Its goal is to find a linear regression model that predicts a set of dependent variables (DV) from a set of independent variables (IV). This is achieved by projecting the IV and DV to new latent variables spaces that have the optimal predictive power. Latent variables, unlike observable variables such as ion channel conductance, are variables that are not observed but are instead inferred from other observables.

PLSR is particularly useful 1) when independent variables are highly correlated (i.e. highly multicollinear) and 2) when the matrix of predictors has more variables than observations. In regression analysis, multicollinearity is problematic in estimating the regression coefficients because some IV effects will be dominated by those of other IVs. Consequently the coefficients will be poorly estimated and any truly significant effect will be lost. Luckily, PLSR does not have this problem and still retains the good predictive power even when working with highly-correlated data sets. Since many variables inside ventricular myocytes may be correlated, PLSR can be considered an ideal alternative to other statistical techniques in analysing parameter-to-output sensitivity. Moreover, since PLSR is reliable in situations where there are more predictors than there are observations, this can save some computational time by reducing the number of jobs needed to be run. PLSR is also useful in situations where one wants to examine the effects of IV on multiple DVs.

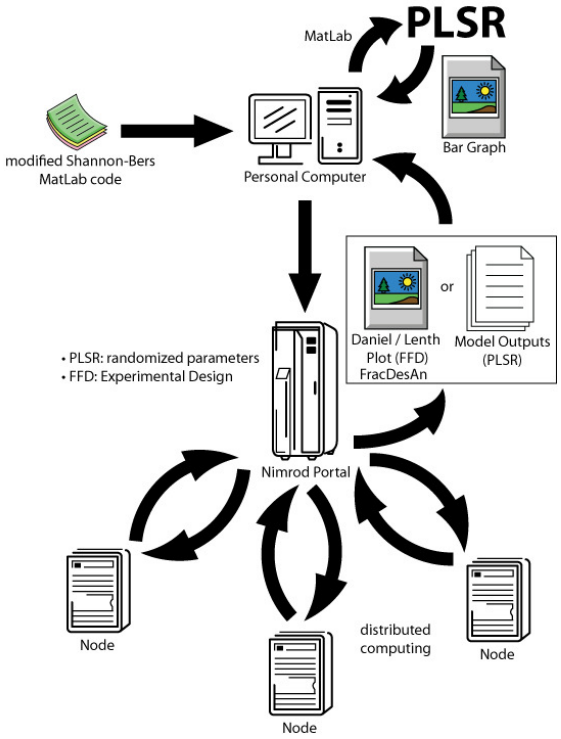
Nimrod/E Fractional Factorial Design

It is often of interest for a researcher to identify variable interactions because they may provide physiological insights into cell dynamics. While collinearity in multiple linear regression is problematic and can lead to inaccurate predictions, Nimrod/E is able to circumvent this problem and successfully identify both main effects and multi-factor interactions. Similar to PLSR, FFD attempts to create a linear model to explain the data generated from the model; it measures and ranks main effects and two-level interactions of input variables with outputs via experimental design. However, the method chooses only those jobs that generate the most significant results by ignoring high-level interactions with little influence, thereby saving a lot of computation time. From the resulting linear model effect coefficients can also be extracted that tells us the correlation between parameters and model outputs.

METHODS

Shannon-Bers (SB) ventricular myocyte model MatLab code was modified to extract the action potential duration (APD) and cytosolic calcium transient ($\Delta[Ca^{2+}]$) as model outputs. The 10^{th} action potential in every simulation was extracted for analysis to ensure that the model is in steady state (steady state usually reached in 7 or 8 stimulations). The model was imported into the Nimrod portal, and fractional factorial design was done by Nimrod/E's experimental design feature to select jobs containing the highest main effect and two-level interactions. The range of parameter values was set at specific standard deviations from the mean. After the design was complete the selected jobs were executed through Nimrod/G parameter sweep. Parameter range was kept small to minimize the occurrence of action potentials that fail to repolarise or otherwise showed grossly unphysiological behaviors. Next, Nimrod outputs were collated into one file to be analysed in *R* and in MatLab. In *R*,

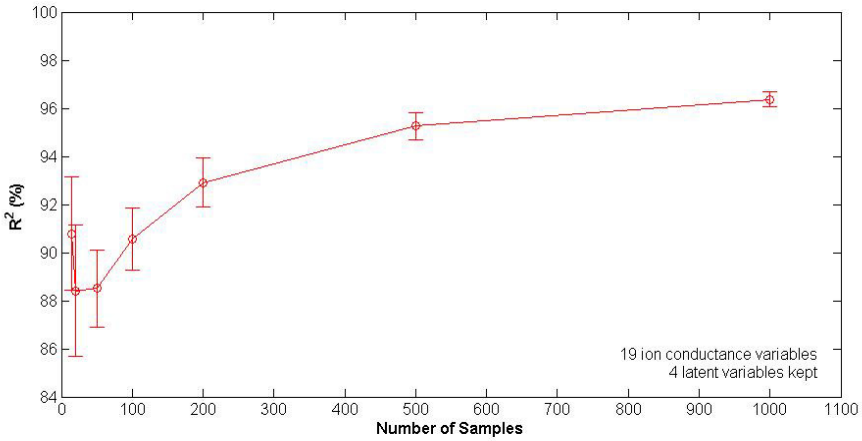
the partial least square (PLS) package (http://cran.r-project.org/web/packages/pls/index.html) was used in *R* for statistical analysis. In MatLab, PLSR algorithm written by Herve Abdi (2004) was used. In both cases, the input (observations x parameter values) and output (observations x model outputs) matrices were normalized (i.e. converted to Z scores) and imported into *R* and MatLab. PLS was performed and the regression coefficients were plotted in bar graphs. The pacing used in all experiments was 1 Hz.



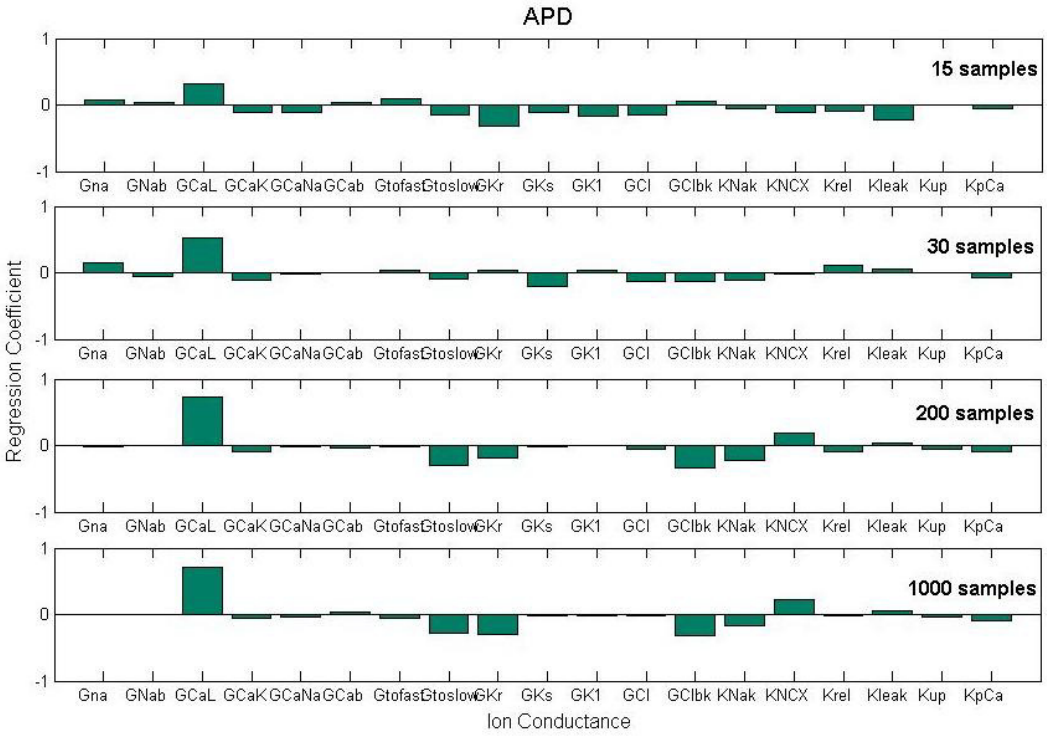
RESULTS

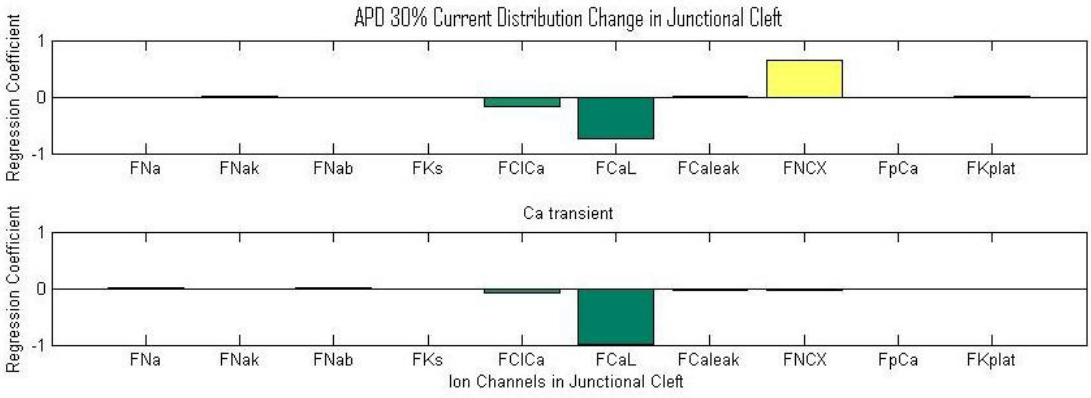
Explanatory capacity (R^2) of the statistical approaches

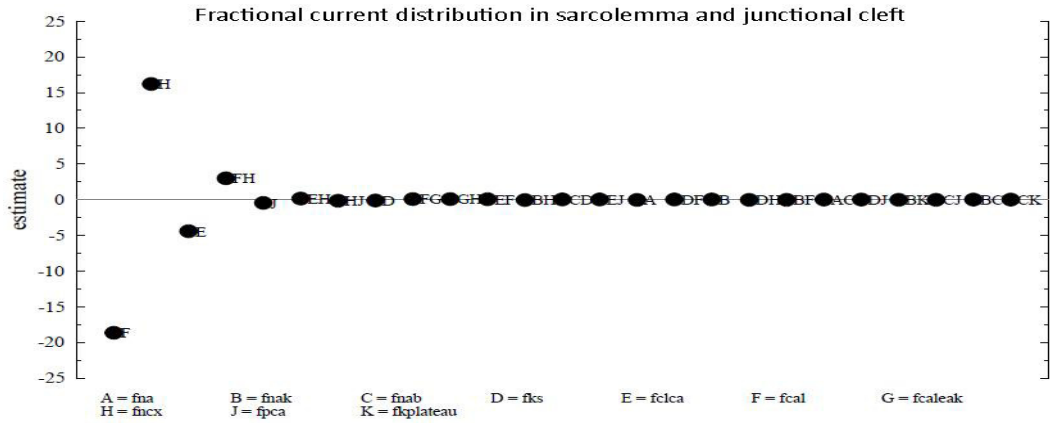
R^2, or the coefficient of determination, quantifies the explanatory capacity of the regression analysis; it indicates the proportion of the variance in the DV that is accounted for by the regression model. An adjusted R-square figure allows a percentage claim, for example if it is 0.45, then it can be said that 45% of the variance is explained by the model. Here, the experiment has 19 conductance variables and 6 outputs. Experimental design first designs and runs 1024 jobs in Nimrod/E. Subsets of the result with size of 15, 20, 50, 100, 200, 500 and 1000 samples are chosen. The respective adjusted R^2 values are calculated and averaged. At the low end of 15 samples, the regression model explains 90.8±2.4% of the variance in the DVs. At the high end of 1000 samples, the model explains 96.4±0.3% of the variance. This relatively low decrease in prediction efficacy demonstrates the robustness of PLSR in analysing parameter sensitivity even in situations where there are more variables than observations.



To further demonstrate this, regression coefficients for ion conductance-APD correlation for different number of samples are plotted. In most cases the regression coefficients are fairly stable in different number of trials.

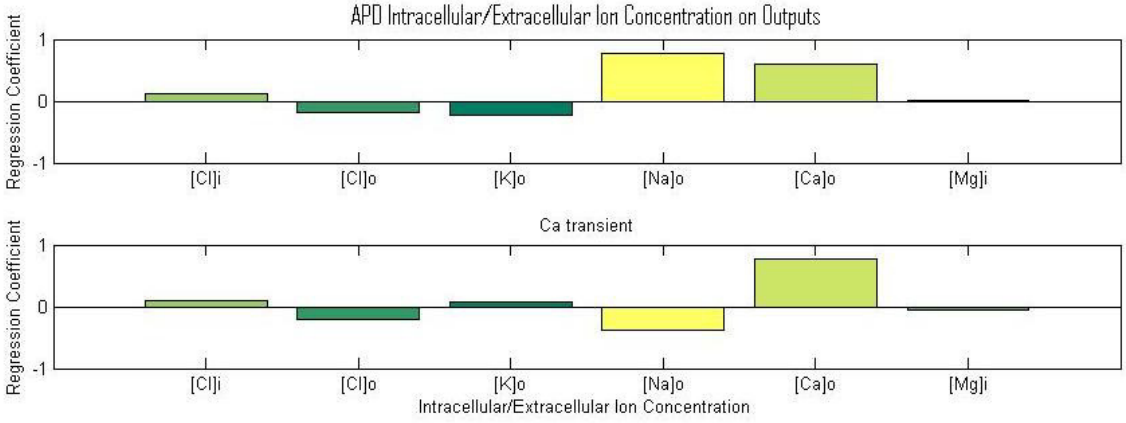


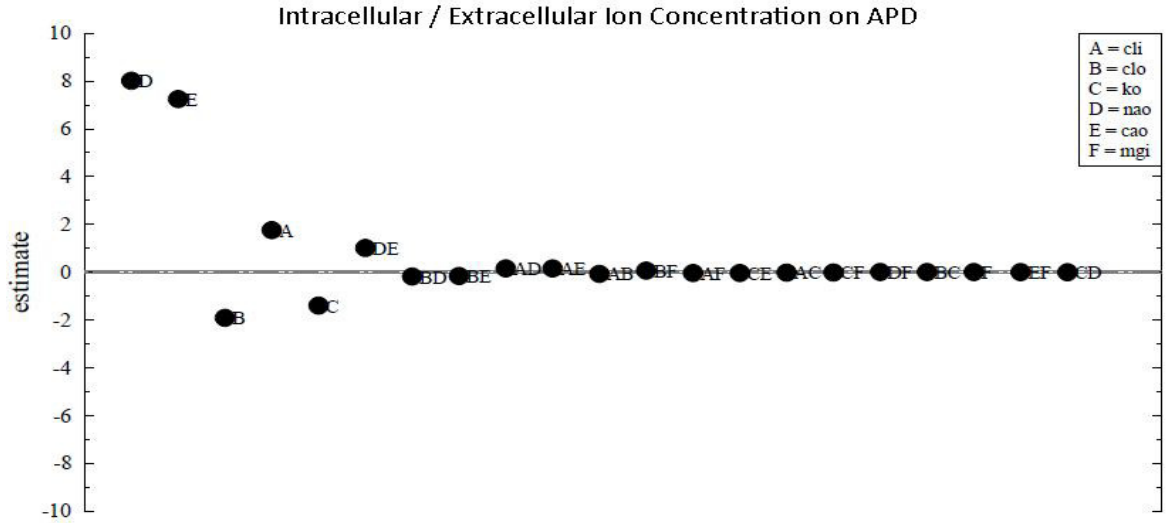




Current distribution in sarcolemma and junctional cleft (model stability testing)

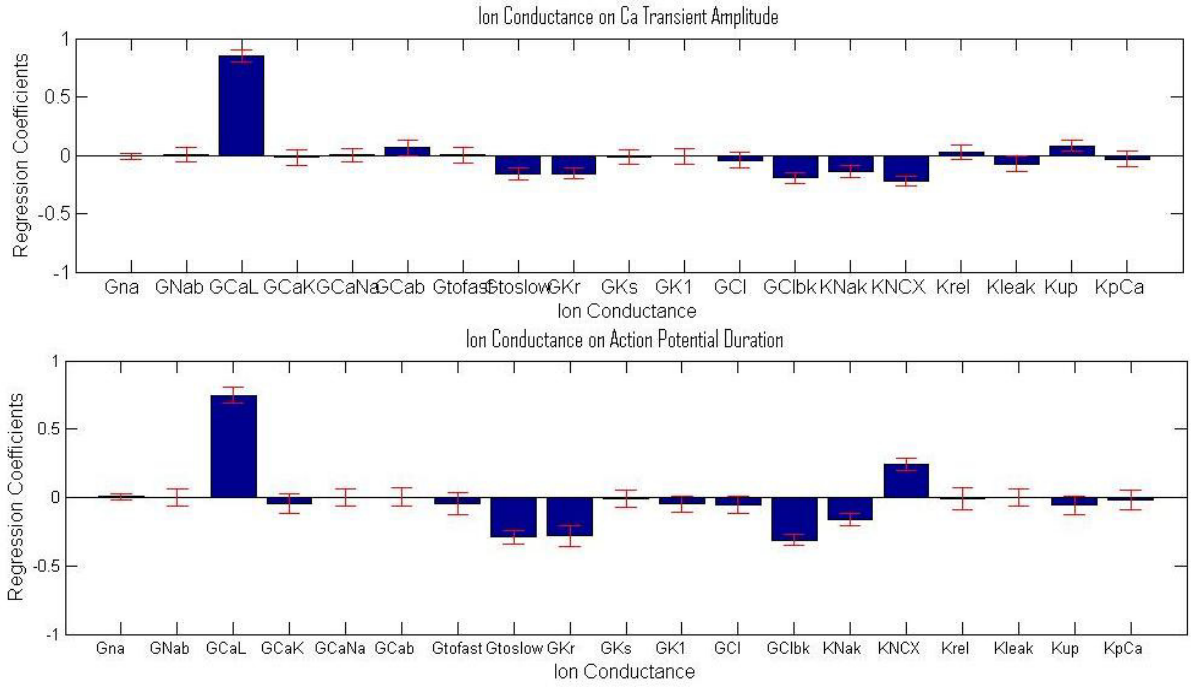
S-B model partitioned the cellular environment into bulk, sarcolemma, junctional cleft and intracellular compartments. Within the sarcolemma and junctional cleft, ion channels are distributed according to a predetermined ratio. Here the graphs show how changes in ion channel densities in the junctional cleft (and consequently the sarcolemma channel densities, since the distribution of channels between the two compartments need to be 100%) will affect model outputs. The result yielded three distributions that significantly affect the outputs. Lowering the densities of L-type Ca channel and Ca-activated Cl channel in the junctional cleft and raising Na-Ca channel distribution in the junctional cleft will increase APD and Ca transient. The underlying physiology is unknown at the moment, but this knowledge can be used to tune model stability and to incorporate effects that reflect real physiology.

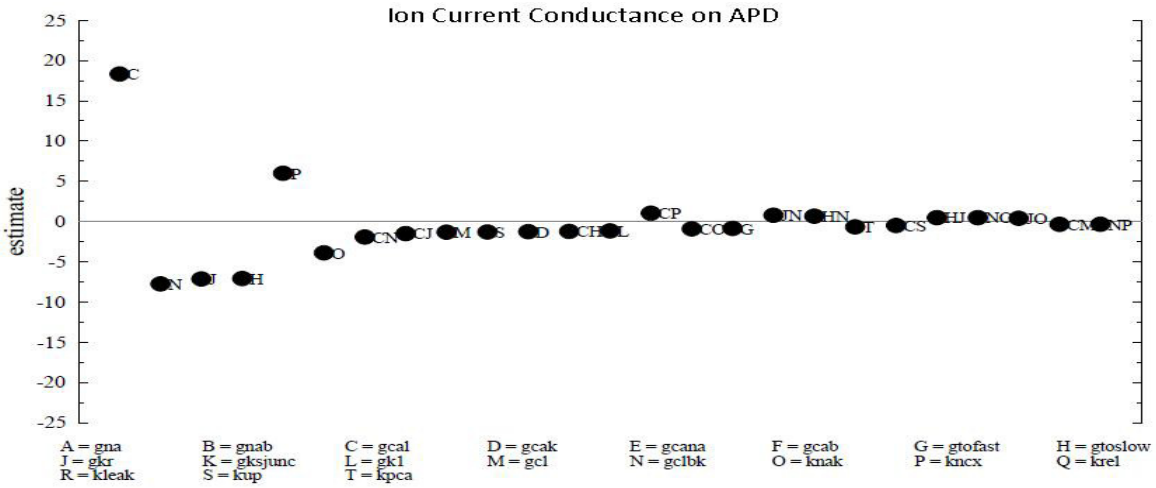




Intracellular / extracellular ion concentration (theory confirmation)

Changes in the ion concentrations between cytoplasm and extracellular space create a concentration gradient that will drive the ion flow between the two spaces. Depending on which ion concentration is changed, APD and Ca transient amplitude may increase or decrease. Here the regression coefficients show how changes in Cl, Ca, K, Mg and Na concentrations inside and outside the myocyte will affect the outputs. Positive coefficient means a rise in a particular ion concentration will increase the magnitude of model outputs. The results indicate that under normal pacing frequency (1 Hz), a rise in extracellular [Na] and [Ca] will increase the APD. Na overload of cardiac cells can accompany various pathologies and induce fatal cardiac arrhythmia. Moreover, a rise in extracellular [Ca] and lowering in [Na] will increase Ca transient. The finding contrasts those of Faber and Rudy (2008); however, more factors – such as pacing frequencies and channel characteristics – need to be considered to accurately evaluate the influence. Therefore, more in-depth study using statistical methods is needed.





Membrane channel ion conductance (model exploration)

The graphs show how different ion conductances affect action potential duration and cytosolic Ca transient amplitude. Positive regression coefficient means increasing the channel conductance will increase the length of excitation, while negative coefficient means increasing the conductance will decrease the duration, etc. The error bars provides a 95% confidence interval for each regression coefficients, so as to validate the prediction. Most results agree with those models Sobie analyzed in his work with the exception of the SR Ca release channels and Na-Ca exchanger. Additionally, background Cl channel, which was a new feature of the S-B model, shows a negative relationship with both outputs. Unlike Sobie's result, SR Ca channels, which have strong implications in facilitating Ca-induced Ca release and moderating Ca transient, seem to play an insignificant role in S-B model. This may be due to a consequence of the model or a flaw in the evaluation process. More work is needed to confirm this. Moreover, in Fox and Kurata's ventricular myocyte model Na-Ca exchanger current negatively affects APD, but it is not the case in S-B model. Overall, most results agree physiologically and are confirmed by experiments. Furthermore, Nimrod/E analysis shows possible interactions between L-type Ca channels (ICaL) and Cl background channel (IClbk) and rapid delayed K channel (IKr).

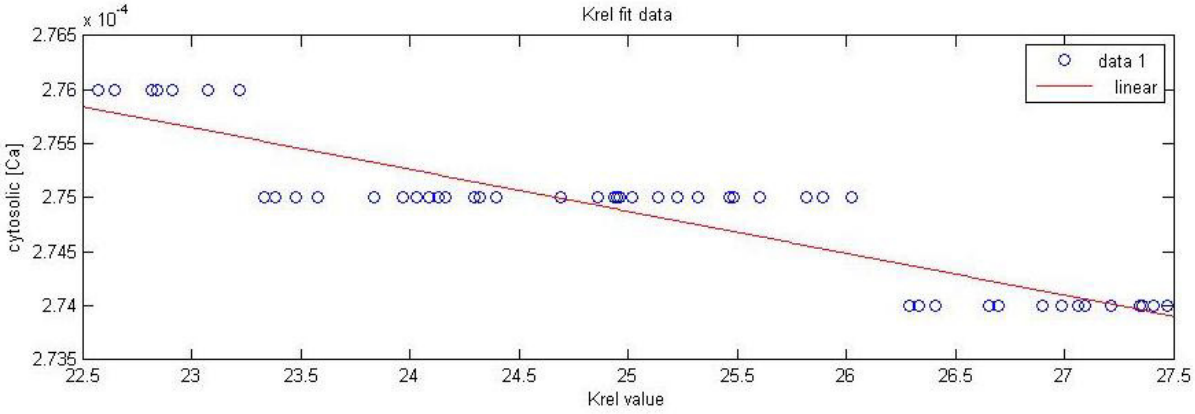
DISCUSSION

In this study I tested the functionality of partial least square regression and fractional factorial design analysis in evaluating parameter sensitivity in Shannon-Bers' electrophysiology model of the ventricular myocyte. In PLSR, inputs were both randomized and chosen specifically using experimental design. In fractional factorial design analysis, parameters were chosen according to a specific algorithm that picks the combination of values that engenders the most significant change in output. The input parameters, along with the generated outputs, were collected as input and output matrices, and the results were subjected to multivariable regression. Input parameters included conductance of ion currents, factors controlling ion-channel gating, intracellular and extracellular ion concentrations, and fractional ion current distribution in sarcolemma and junctional cleft. Outputs included physiologically significant measures such as action potential duration and cytosolic calcium transient amplitude.

Individual Conductance's Effect on cytosolic [Ca]

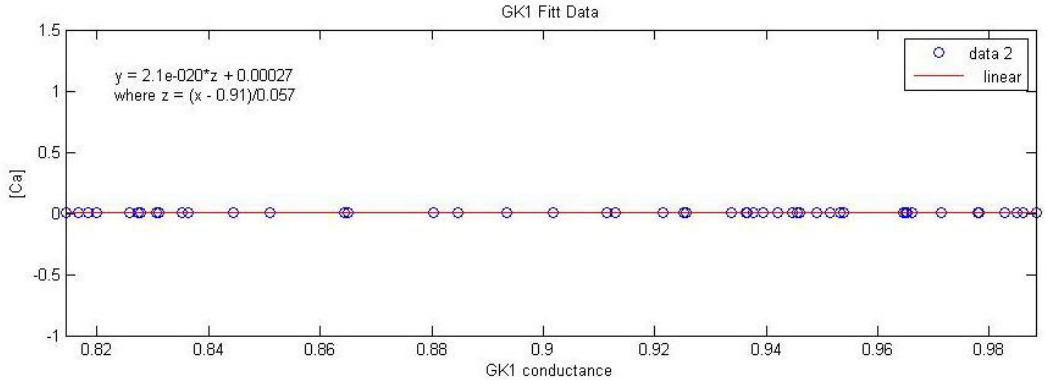
After identifying the 8 relatively significant ion channel conductance variables, I ran experiments where I randomly varied one variable individually and collected 50 samples for

fitting using linear regressions. I first plotted the conductance values vs. respective outputs and fit the data points to a line. I thought it would be interesting to examine the relationship of the conductance to APD and [Ca]. For instance, if there is a logarithmic or sigmoidal relationship between a channel conductance and APD, then this kind of information could shed some light on the mechanism of ion flow, etc. While linear regression on channel conductance vs. APD didn't yield a noticeable general trend, the regression on [Ca] vs. APD yielded some pretty interesting finds.

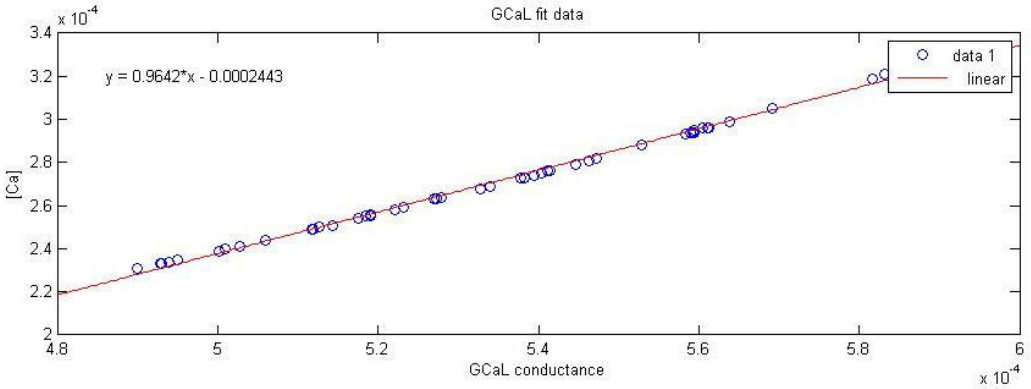


Kup's fit data also shares the same downward trend as Krel. This stepwise inverse relationship seems to be a feature of the Shannon-Bers model. Since both of these channels are located in the SR, this nonlinear relationship might be related to the channel's intrinsic opening probability. This stepwise relationship for Kup and Krel to [Ca] may explain why the regression coefficients for these two conductances are low even though there is an obvious trend between Ca transient and Krel and Kup values. Since the relationship is not linear, PLS (a "linear" regression method) will do a poor job of predicting the variable and might consider it as insignificant.

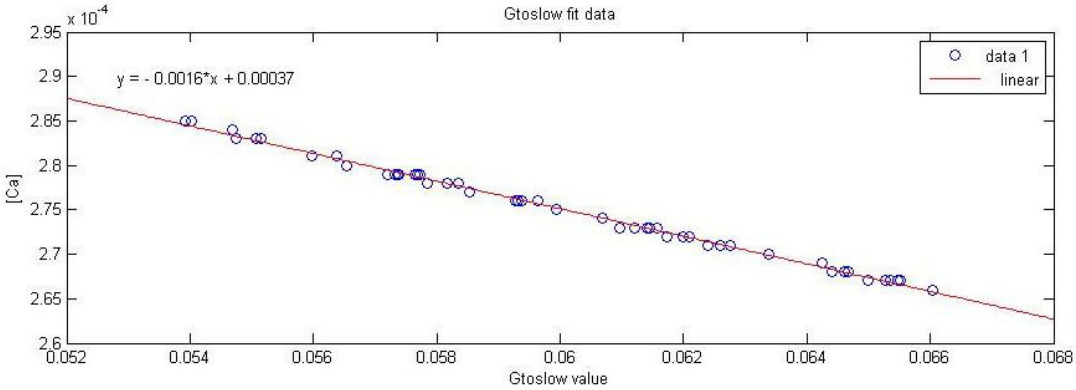
A nontrivial implication from this graph is that as the Ca-induced Ca release channel (CICR) conductance increases, cytosolic [Ca] seems to decrease. While CICR channel is located in the junctional cleft between the SR and the SL (the fit data for that shows a direct relationship), this stepwise relationship is still counterintuitive because CICR channel opening should increase free cytosolic [Ca]. Perhaps other mechanisms are involved that can explain this trend.



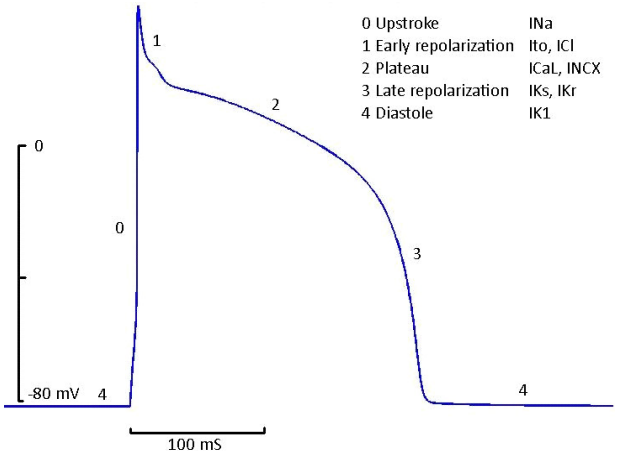
The fit data suggests that GK1 has no influence on the [Ca], which makes sense physiologically since the K rectifier channel does not conduct Ca ions.



As expected, the opening of L-type Ca channel will lead to an increase in cytosolic [Ca].



KNaK, GClbk and GKr all share the same downward trend of inversely affecting the cytosolic [Ca].



Ion Channels-APD correlation confirmation

Upon a sudden change in membrane potential that reaches suprathreshold, the voltage-gated Na channel (INa) opens and allows an influx of Na into the cell that further depolarizes the cell. Subsequently voltage-gated transient outward K channels (Itoslow) will trigger an efflux of K that will repolarise the membrane potential. K channel controls early repolarization phase 1, so down-regulation of this channel may contribute to action potential prolongation and arrhythmia in congestive heart failure (Kaab et al., 1998). This correlation is reflected in the regression coefficients by Gtoslow's negative effect on APD. Next, inward Ca currents carried by L-type Ca channels contributes to the phase 2 plateau of AP. The opening of this channel leads to the influx of Ca into the junctional cleft, which will causes Ca binding to Ryanodine receptor (RyR) and trigger Ca-induced Ca release (CICR) from the sarcoplasmic reticulum (SR). Reduction in amplitude of ICaL may be responsible for action potential shortening and atrial fibrillation (Le Grand et al., 1994; Ouadid et al., 1995). Once again the regression coefficient shows a positive relationship between ICaL and APD and supports this finding. Furthermore, Reduced channel inactivation results in enhanced depolarizing Ca current and lead to early afterdepolarization, a factor that increase the risk for arrhythmia (Splawski et al., 2005). The amount of Ca entering the junctional cleft through the SL Ca channel determines the extent of RyR activation. Since RyR are located throughout the junction cleft of the SR membrane, multiple Ca bindings need to occur to activate RyR and trigger the SR Ca release. The influx of large amounts of Ca into the cytoplasm via SR Ca channel delays the repolarization step and is responsible for the plateau phase of cardiac AP and directly determines the APD. As the opening of SR Ca channel depletes the store of SR Ca, SR Ca-ATPase will pump free systolic Ca into the SR and reload the SR for the next episode of CICR. Therefore, a good balance between SR Ca release and uptake is needed to maintain a normal AP. In contention with membrane depolarization during the plateau phase is an efflux of repolarizing K current caused by rapid delayed rectifier K channel (IKr). The balance of the two ion fluxes determines the duration of AP. Another K channel activated during the plateau phase is the inward rectifying K channel, which conducts K depending upon the membrane potential; at membrane potentials negative to K's reversal potential (~-100 mV), the channel conducts K, and at membrane potentials positive to K's reversal potential, the channel conducts K out of the cell and repolarises the cell. Thus, during an AP when the membrane potential is positive there will be an influx of K into the cytoplasm. Essentially it will extend the plateau phase and prolong APD. Additionally, the regression coefficients also suggest that SL background Cl channel can negatively affect the APD. This is probably due to the influx of Cl during the repolarization phase which shortens the plateau phase. Na-K and Na-Ca

exchanger seem to play a minor role in influencing APD; however, the direction of influence, as is suggested by the regression coefficients, is not reliable. Several experiments have implicated the role of Na-Ca exchanger and Na-K in initiating CICR. For instance, Na-Ca exchanger has the appropriate Km to back up the back up the Ca reaccumulation in the SR soon after the peak influx of myoplasmic [Ca]i; however, since the amount of Ca and Na conducted through the two channels are relatively small compare to L-type Ca channel, they do not play a significant role in triggering SR Ca release.

FUTURE WORK

The next thing to do to realize the versatility of PLSR in studying cardiac models is to evaluate the validity of results. This includes examining the explanatory capacity (R^2) of PLSR in comparison to other regression methods such as multiple regression and principal component analysis, and also examining the results' significance (reflected in p-values and coefficients) in different parameter variations. Explanatory capacity is paramount in evaluating the features of cardiac models, for the type, sign and consistency of relationship between DVs and IVs are the basis for establishing fundamental links between physiological meaning and model parameters.

ACKNOWLEDGEMENT

This work was supported by National Science Foundation grant IOSE-0710726. I would like to give special thanks to Tom Peachey, Neil Diamond, Blair Bethwaite, Colin Enticott, David Abramson (Monash University) and Anushka Michailova (UCSD) for discussion and technical support. Additionally, I would like to thank Gabriele Wienhausen, Peter Arzberger, Teri Simas and Rob Gray for making PRIME a dream come true for me. ☺

REFERENCE

1. Abdi, H. 2010. Partial least square and projection of latent structure regression (PLS regression). www.wiley.com/wires/compstats
2. Fabiato, A. Calcium-induced release of Calcium from the cardiac sarcoplamsic reticulum. 1983. Am J Physiol Cel Physiol 245: C1-C14
3. Geladi, P., Kowalski, B. 1986. Partial least-squares regression – a tutorial. – Anal. Chim. Acta. 185: 1-17
4. Haenlein, M., Kaplan, A. 2004. A Beginner's Guide to Partial Least Square Analysis. Understanding Statistics, 3(4), 283-297
5. Henk E., Penelope B. 2007. Calcium and Arrhythmogenesis. Physiol. Rev. 87: 457-506
6. Kirsch, G. 1999. Ion Channel Defects in Cardiac Arrhythmia. J. Membrane Biol. 170, 181-190
7. Mevik, B., Wehrens, R. 2007. The pls package: Principal Component and Partial Least Squares Regression in R. J of Stat. Software. Vol. 18, Issue 2. http://jstatsoft.org/
8. Peachey, T., Sudholt, W., Michailova, A. 2008.Fractional Factorial Design for Parameter Sweep Experiments using Nimrod/E.
9. Puglisi, J., Bers, D. 2001. LabHEART: an interactive computer model of rabbit ventricular myocyte ion channels and Ca transport. Am J Physiol Cell Physiol 281: C2049-C2060
10. Shannon, T., Bers, D. 2004. A Mathematical Treatment of Integrated Ca Dynamics within the Ventricular Myocyte. Bio. Phys. J 87(5) 3351-3371
11. Sobie, E. 2008. Parameter Sensitivity Analysis in Electrophysiological Models Using

Multivariable Regression. Bio. Phys. J 96(4) 1264-1274