

# Parameter Sensitivity Analysis of Cardiac Purkinje fiber Single Cell Model Using Partial Least Square Regression

Durga Choudhary<sup>1</sup> Dr. Akhil Ranjan Garg<sup>2</sup>

<sup>1</sup>Research Scholar

<sup>1,2</sup>Department of Electrical Engineering

<sup>1,2</sup>M.B.M Engineering College, J.N.V.U, Jodhpur, India

**Abstract**— In traditional method of parameter sensitivity analysis only one variable can be varied, and then impact on the output is observed, but the correlation between input variables and its effect on output was not observed. To overcome this problem the new technique was developed by Wold et al. In this technique we can analyze or predict a set of dependent variables from a set of independent variables (predictors). This prediction is done by a set of orthogonal factors (latent variables) obtained from a set of predictors and these latent variables have better predictive power. It combines the features of PCA and multiple linear regression. The PLS regression technique is used to determine the strength of the impact of multiple independent variables on a dependent variable using regression coefficient matrix which is a reflection of parameter sensitivities of input parameters. In this work we have done parameter sensitivity analysis of the already existing cardiac single cell Purkinje fiber (Noble) model.

**Keywords:** ionic conductance, injected current, regression coefficient, residual, standard deviation

## I. INTRODUCTION

PLS is a good alternative to the more classical multiple linear regression and principal component regression methods because it is robust. PLS is designed to reflect theoretical and empirical conditions present in the behavioral and social sciences, where there is a lack of solid theories and scarce knowledge. This kind of modeling is called soft modeling (Wold 1980). Mathematical and statistical procedures underlying the system are rigorous and robust (Wold 1979), however mathematical model is soft in the sense that it makes no measurement, distribution and sample size assumptions. The goal achieved is milder than hard modeling.

Parameter sensitivity analysis and consideration of the variability can provide insight into a number of important issues in cardiac physiology. The insights gained from such studies have important implications not only for building better mathematical models, but possibly also for the diagnosis and treatment of diseases. Improved methods for translating preclinical measurements into patient-specific models would substantially improve the ability to create models of a broader range of complex pathologies and robustly and reliably use this information to inform clinical decisions [1]. In their work they introduced an effective method to model empirical action potentials of specific patients in the human atria based on the minimal model of Bueno-Orovio, Cherry and Fenton adapted to atrial electrophysiology. In recent years, the parameter sensitivity analysis is done by varying all parameter at once, to capture and understand the intrinsic variability of populations [2]. In the various models of cardiac cells [3,4 and 5] done the parameter sensitivity analysis using traditional method.

In this work, we will discuss how PLS is useful for the researcher in diverse fields to understand and predict the causes of functional variability between different samples. Here PLS is used to constrain free parameters in single cell Purkinje fiber (Noble) model [6]. Noble model was self-oscillatory model, without external current the action potential was generated. The cell membrane potential or voltage  $V$  of Noble model which follows an ordinary differential equation shown below.

$$\frac{dV}{dt} = \frac{-1}{C_m} [(\bar{g}_{Na}m^3h + \bar{g}_C)(V - E_{Na}) + (\bar{g}_{K1} + \bar{g}_{K2}n^4)(V - E_K) + \bar{g}_L(V - E_L)] \quad (1.1)$$

The goal of the PLS regression is to generate a new simplified and empirical model which predicts output resulting from a new set of input parameters. It also generates the regression coefficient matrix which is a reflection of parameter sensitivities of input parameters. Computations are performed with the randomized input parameters and the AP characteristics computed from the simulations are recorded as corresponding output matrices. Then the results are subjected to multivariable regression.

## II. PLS REGRESSION FOR NOBLE MODEL

In Noble model, there are some dependent parameters and some independent parameters. The maximal values of ionic conductances  $\bar{g}_{Na}$ ,  $\bar{g}_C$  and  $\bar{g}_{K2}$  are independent parameters which are varied randomly by scaling the standard maximal conductance values listed in published model[2]. The probability distribution in Matlab is done using the function `makedist`. This function creates the log-normal distribution of object with a mean ( $\mu$ ) and standard deviation ( $\sigma$ ). The probability that a conductance is doubled equalled the probability that its value is halved. Using this log-normal distribution, we ensured that the conditions of a negative conductance never encounter. The parameter ' $\sigma$ ' specifies the standard deviation of the distribution of log transformed variables, controlled the extent to which parameter is varied. Before performing PLS regression, data are normalized using a Z-score method with a mean of zero and standard deviation of one.

### A. Methodology for sample collection

For simulation, computations are performed for action potential of fifth pulse and the successive diastolic interval pair. The program was run for a 6000msec time period. For each set of chosen parameters the corresponding action potential was computed. This procedure was repeated numerous times to generate input and output matrices. The simulation is performed for each sample till the shape of the action potential is proper or peak overshoot voltage becomes negative. In this process 71.6% samples are rejected due to grossly unphysiological behavior, such as APs that never repolarized. The program is run for 2500 samples, but 29.4%

samples have been accepted, so  $n=735$ . Simulations are performed at  $\sigma=0.15$ , so 95% ( $\pm 2$  standard deviation) of the randomly chosen maximal conductances will vary 35% ( $e^{2(0.15)}$ ) above the standard value and 26% ( $e^{-2(0.15)}$ ) below standard value.

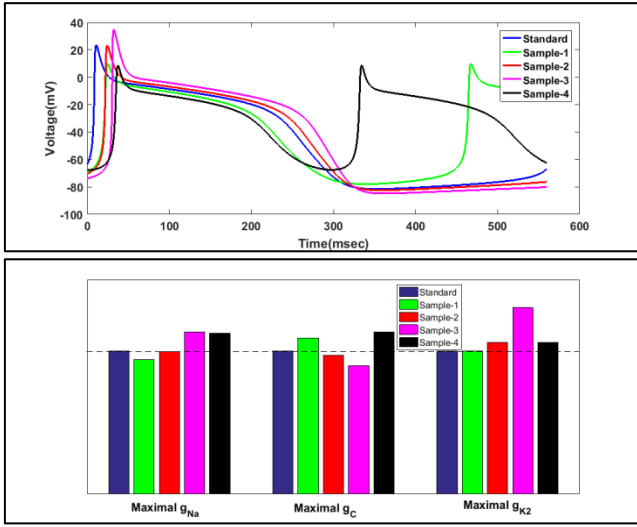


Fig 1: (a) Changes in AP morphology, compared with standard AP (blue colour), produced by randomly varying ionic conductances in the Noble model (b) Bar chart showing the corresponding values of the AP with the same colour coding.

The standard value for  $\overline{g_{Na}}$ ,  $\overline{g_C}$  and  $\overline{g_{K2}}$  are  $4\text{mS/mm}^2$ ,  $0.0014\text{mS/mm}^2$  and  $0.012\text{mS/mm}^2$  respectively.

Input matrix [X] consists of independent parameters

$$[X] = [\overline{g_{Na}} \quad \overline{g_C} \quad \overline{g_{K2}}]_{n \times p}$$

Where, ( $n=735$ ) is the number of samples or trials of random parameters and ( $p=3$ ) is the number of independent parameters varied in the model.

Output matrix[Y] consists of dependent parameters

$$[Y] = [\text{APD}, V_{\text{rest}}, V_{\text{peak}}, \text{BCL}]_{n \times m}$$

Where, ( $m=4$ ) is number of dependent parameters in the model. The parameter APD is action potential duration at 90% of repolarization.  $V_{\text{rest}}$  is resting potential.  $V_{\text{peak}}$  is peak voltage and BCL is basic cycle length. The size of X matrix is  $735 \times 3$  and the size of Y matrix is  $735 \times 4$ .

Figure 1 shows action potentials produced by randomly varying the  $\overline{g_{Na}}$ ,  $\overline{g_C}$  and  $\overline{g_{K2}}$  in the Noble model for 4 samples. The action potential generated for the corresponding first four values of the X matrix along with the standard value of the Noble model is shown in the figure 1(a). Normalized values of  $\overline{g_{Na}}$ ,  $\overline{g_C}$  and  $\overline{g_{K2}}$  are given in the bar graph, color-coded to the corresponding AP traces as shown in figure 1(a). These random changes could cause APD,  $V_{\text{rest}}$ ,  $V_{\text{peak}}$  and BCL to increase or decrease, depending on the combination of conductances in a particular trial. The dotted line in figure 1(b) specified standard values of  $\overline{g_{Na}}$ ,  $\overline{g_C}$  and  $\overline{g_{K2}}$  respectively. The value of  $\overline{g_C}$  and  $\overline{g_{K2}}$  is very less as compared to  $\overline{g_{Na}}$ , so by scaling of standard values of  $\overline{g_C}$  and  $\overline{g_{K2}}$  same level is obtained.

### B. Z-score normalization

Random-scale factors obeyed a log-normal distribution, so, most values in the input matrices were log-transformed before computing the Z-score normalization. Distributions of

APD and BCL are right-skewed, so these values are also log-transformed, and retransformed back into the original units for the display of results. By visual inspection,  $V_{\text{rest}}$  and  $V_{\text{peak}}$  are approximately normally distributed; therefore, these variables are not log transformed before Z-score normalization. Before performing regression, data is pre-processed to suitable form, so that it will be suitable for regression analysis. Normalization of data is done using a Z-score method having a mean of zero and standard deviation of one. The normalized inputs (X and Y matrix) are regressed using the NIPALS algorithm [7].

### C. Regression coefficient matrix $B_{PLS}$

The regression coefficients in the  $B_{PLS}$  matrix indicate the impact of input parameter on the output, 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.

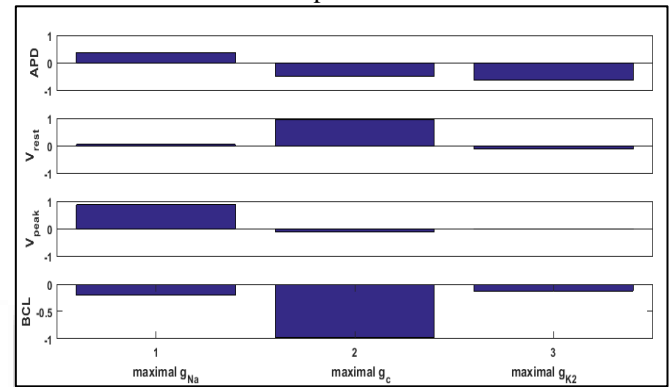


Fig. 2: Columns of regression coefficient matrix  $B_{PLS}$  produced by NIPALS algorithm

The values in  $B_{PLS}$  are plotted in the form of bar chart can be understood quantitatively as follows. Because the input and output matrices are mean-centred and normalized to standard deviations computed column-by-column, each element of  $B_{PLS}$  matrix 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.

To observe the effect of independent parameters on individual dependent parameters, in the form of a graph, the bar chart is plotted from the  $B_{PLS}$  matrix which is shown in figure 2. The first row in bar chart represents the impact on APD by various maximal conductances. Similarly, second, third and fourth row represents the impact on  $V_{\text{rest}}$ ,  $V_{\text{peak}}$  and BCL due to various maximal conductances. From figure 2, it has been observed that,

- 1) Increase in  $\overline{g_{Na}}$ , and decrease in  $\overline{g_C}$  or  $\overline{g_{K2}}$ , lengthen the APD,
- 2)  $V_{\text{rest}}$  is determined by a balance between  $\overline{g_{Na}}$  and  $\overline{g_{K2}}$ . Whereas,  $\overline{g_C}$  has major impact on  $V_{\text{rest}}$  and positively correlated with it.
- 3)  $\overline{g_{Na}}$  is the only conductance exerting a substantial effect on  $V_{\text{peak}}$ .
- 4) All the maximal conductances are negatively correlated with BCL, but it is strongly affected by the maximal conductance  $\overline{g_C}$ .

D. Predictions of dependent parameters in PLS regression

In PLS regression the  $B_{PLS}$  matrix is used to predict the output using the following equation

$$\hat{Y} = X * B_{PLS} \quad (1.2)$$

Where,  $\hat{Y}$  is predicted output, which is close to the original output matrix 'Y'?

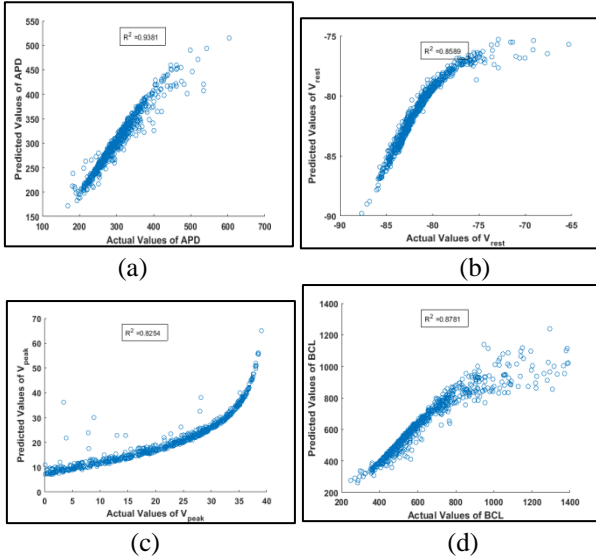


Fig. 3: Scatterplots of (a)APD (b) $V_{rest}$  (c) $V_{peak}$  (d)BCL

Independent variables in PLS regression are called predictors. Figure 3 shows the predictions of dependent variables which are predicted by the linear empirical model generated by PLS regression. To observe these predictions in the graphical form, the scatterplots are plotted for APD (figure 3a),  $V_{rest}$  (figure 3b),  $V_{peak}$  (figure 3c) and BCL(figure 3d). For each of the four outputs, the plot shows the ‘actual’ value, calculated by the numerical integration of the Noble differential equations, versus the prediction of the empirical regression model. The residual  $R^2$  is the sum of squares of the difference between the actual and predicted responses. The large residual ( $R^2$ ) values (0.8254–0.9381) indicate that the simplified linear model is highly predictive, despite the nonlinear differential equations in the Noble model. The residual  $R^2$  indicates percentage of data which are correctly predicted. The standard value of APD is 284.7msec. The figure 3(a) consists of actual and predicted values are predicted correctly the value of APD with the residual  $R^2$  of 0.9381. This residual shows that almost 93.81% of the values are predicted correctly. The figure 3(b) shows the actual and predicted values of the resting potential. The standard value of resting potential is -81.5789mV. The residual  $R^2$  of the resting potential is 0.8589, this shows that 85.89% of the resting potential values are predicted correctly. Similarly, figure 3(c) shows actual and predicted values of the action potential peak voltage. The standard value of action potential peak voltage is 23.368mV. The residual  $R^2$  of the peak potential is 0.8254; which shows that 82.54% of the resting potential values are predicted correctly. Further figure 3(d) shows actual and predicted values of the basic cycle length (BCL). The standard value of BCL is 564.1msec. The residual  $R^2$  of the BCL is 0.8781; this shows that 87.81% of the BCL values are predicted correctly.

E. Effect of sample selection on regression coefficient matrix ( $B_{PLS}$ ) and residual ( $R^2$ )

In this subsection we will examine the impact of a number of samples on the regression coefficient matrix ( $B_{PLS}$ ). We have executed the PLS regression for various samples. For the random simulation with a standard deviation of 0.15, for 100 samples only 48 samples are accepted. Likewise for 240 samples only 106 samples are accepted, for 1200 samples we got 505 samples. Also for the same deviation we got 1000 and 2088 data from 2500 and 5100 samples respectively. The resultant  $B_{PLS}$  matrix of these data is shown in the figure 4. The figure 4 consists of four subplots, each showing the parameters of the output matrix. Each subplot consists of five bars each showing the value of the regression coefficient for the number of samples. The figure 4 is the replication of figure 2 with each bar for different number of samples. It can be seen from figure 4, that even if the number of samples are less ( $n=48$ ), PLS still provides reasonable estimates of regression coefficients.

As discussed earlier  $\bar{g}_{K2}$  has major impact on APD, which has 5.06% of variation as we increase the number of samples. Similarly  $\bar{g}_C$  has main impact on  $V_{rest}$  and BCL, so it has 2.64% and 1.52% of variation respectively. Whereas,  $\bar{g}_{Na}$  has major impact on  $V_{peak}$  with 3.63% of the variation. For analysis of computational models, of course, excellent estimates of regression coefficients can always be obtained by running additional simulations to generate more samples.

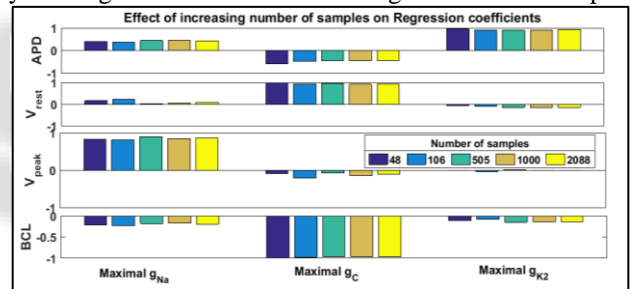


Fig. 4: Effect of increasing number of samples on  $B_{PLS}$  matrix

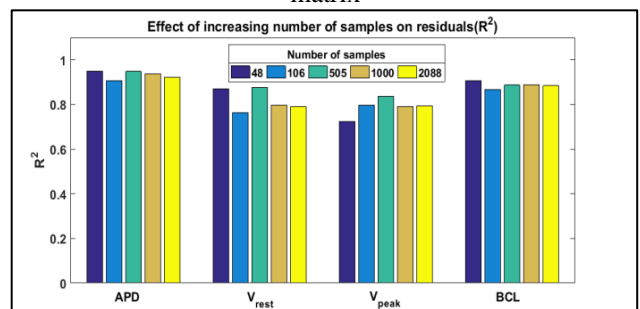


Fig. 5: Effect of increasing number of samples on residual ( $R^2$ )

Further to observe the impact of number of samples on residuals bar chart is plotted for  $R^2$  for dependent parameters APD,  $V_{rest}$ ,  $V_{peak}$  and BCL. It can be observed from figure (7) that maximum value of  $R^2$ (for APD and BCL) is obtained for 48 samples. Likewise, maximum value of  $R^2$ (for  $V_{rest}$  and  $V_{peak}$ ) is obtained for 505 samples. The program is run for random trials, so these results will not be fixed, but variation in the value of  $R^2$  is less.

F. Effect of sigma ( $\sigma$ ) on regression coefficient matrix ( $B_{PLS}$ ) and residual ( $R^2$ )

Thus far, simulations shown in figure 1-5, input parameters are varied over a relatively tapered range. The value of the parameter  $\sigma$ , which controls random variation of maximum input parameters, was set at 0.15. In biological samples, however, the expression levels of ion channels can sometimes change more dramatically. In order to observe the effect of increasing sigma ( $\sigma$ ) on  $B_{PLS}$  matrix and residual ( $R^2$ ), we have generated the random generations for input parameters for various values of sigma ( $\sigma$ ) which are 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3. Larger values resulted in grossly unphysiological behavior, such as APs that never repolarized, in a significant proportion of trials.

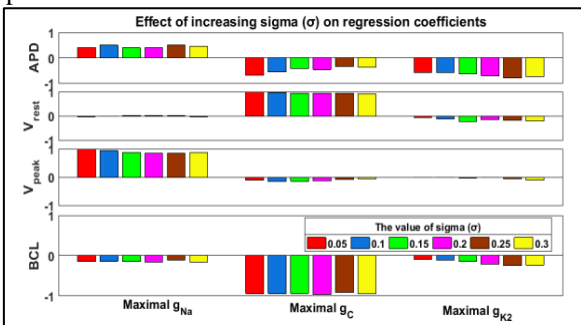


Fig. 6: Effect of increasing sigma ( $\sigma$ ) on  $B_{PLS}$  matrix

The figure 6 is the replication of figure 2 with each bar for different value of sigma ( $\sigma$ ). It can be seen from figure 6, that even if the value of sigma is more ( $\sigma=0.3$ ), PLS still provides reasonable estimates of regression coefficients.

As discussed earlier  $\overline{g_{K2}}$  has major impact on APD, which has 13% of variation as we increase the value of sigma ( $\sigma$ ). Similarly  $\overline{g_C}$  has main impact on  $V_{rest}$  and BCL, so it has 2.78% and 1.56% of variation respectively. Whereas,  $\overline{g_{Na}}$  has major impact on  $V_{peak}$  with 6.2% of variation.

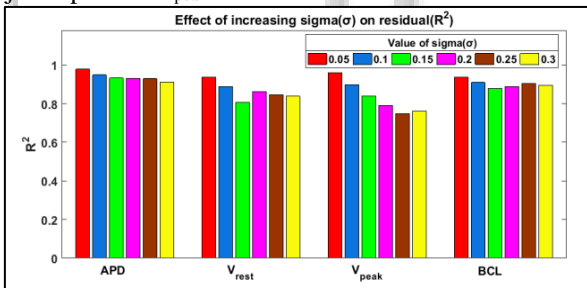


Fig. 7: Effect of increasing sigma ( $\sigma$ ) on residual ( $R^2$ )

Figure 7 shows that the predictive power of the regression model declined somewhat with an increase in  $\sigma$ . It can be observed from figure 7,  $R^2$  is decreasing from 0.97 to 0.91 for APD, whereas, 0.93 to 0.89 for BCL. Similarly,  $R^2$  is decreasing from 0.93 to 0.83 for  $V_{rest}$ , but for  $V_{peak}$   $R^2$  is decreasing from 0.96 to 0.76. 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.

III. CONCLUSION

PLS regression illustrates the potential utility of multivariable regression for parameter sensitivity analysis. 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 PLS regression responded in the same manner, even when the number of samples are small. This is the main advantage of this PLS regression that the computational cost to do parameter sensitivities can be assessed by performing only a few hundred simulations with randomly varying parameters.

REFERENCES

- [1] Yvonne Richter, Pedro G. Lind, Philipp Maass, 2018. Modeling specific action potentials in the human atria based on a minimal single-cell model, Editor: Elena Tolkacheva, University of Minnesota.
- [2] Sobie EA. Parameter sensitivity analysis in electrophysiological models using multivariable regression. Biophys J. 2009;96:1264–1274.
- [3] Corrias A, Giles W, Rodriguez B. Ionic mechanisms of electrophysiological properties and repolarization abnormalities in rabbit Purkinje fibers. Am J Physiol Heart Circ Physiol. 2011;300:H1806–H1813..
- [4] Nygren A, Fiset C, Firek L, Clark JW, Lindblad DS, Clark RB, Giles WR. Mathematical model of an adult human atrial cell: the role of  $K^+$  currents in repolarization. Circ Res. 1998;82:63–81.
- [5] Romero L, Carbonell B, Trenor B, Rodriguez B, Saiz J, Ferrero JM. Systematic characterization of the ionic basis of rabbit cellular electrophysiology using two ventricular models. Prog Biophys Mol Biol. 2011;107:60–73.
- [6] NOBLE, 1961. A modification of the hodgkin-huxley equations applicable to purkinje fibre action and pacemaker potentials, J. Phy8iol. (1962), 160, pp. 317-352
- [7] 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.