Gray Box Model with an SVM to Represent the Influence of $PaCO_2$ on the Cerebral Blood Flow Autoregulation

Max Chacón[1], Mariela Severino[1], and Ronney Panerai[2]

[1] Departamento de Ingeniería Informática, Universidad de Santiago de Chile,
Av. Ecuador 3659,
Casilla 10233, Santiago, Chile
max.chacon@usach.cl, mariela.severino@gmail.com
[2] Medical Physics Group, Department of Cardiovascular Sciences, Faculty of Medicine,
University of Leicester, Leicester LE1 5WW, UK
rp9@le.ac.uk

Abstract. Since the appearance of methods based on machine learning, they have been presented as an alternative to classical phenomenological modeling and there are few initiatives that attempt to integrate them. This paper presents a hybrid paradigm called *gray box* that blends a phenomenological description (differential equation) and a Support Vector Machine (SVM) to model a relevant problem in the field of cerebral hemodynamic. The results show that with this type of paradigm it is possible to exceed the results obtained with phenomenological models and also with the models based on learning, in addition to contributing to the description of the modelled phenomenon.

Keywords: Gray Box Model, Support Vector Machine, Cerebral hemodynamic, $PaCO_2$.

1 Introduction

In many fields of science and engineering one has partial knowledge of the phenomenon that it is desired to model, having available some equation that describes it partially (deterministic model) –this problem appears more often when the phenomenon is nonlinear–. On the other hand, there are methods based on learning, such as artificial neural networks (random model), which allow modeling nonlinear phenomena but do not describe adequately the principles of the phenomena.

In the 1990s, Psichogios and Ungar [1] proposed the methods called *gray box*, which involve a hybrid strategy that mixes the phenomenological knowledge of an equation, usually differential (*white box*), and an automatic learning method like neural networks (*black box*) to make a more accurate description of the phenomenon (*gray box*). Thompson and Kramer [2] classified these methods into basically two structures. The so-called "*series configuration*", where the neural network participates by adjusting parameters of the differential equation with the purpose of incorporating the data variations in the deterministic model, and the other alternative is the "*parallel configuration*", where the neural network adjusts the results of the differential equation

C. San Martin and S.-W. Kim (Eds.): CIARP 2011, LNCS 7042, pp. 630–637, 2011.
© Springer-Verlag Berlin Heidelberg 2011

to the data, estimating the residues between both. Representative applications of these methods are found in the fields of chemistry and bioprocesses [3-5].

In this paper we propose the use of the *gray box* method, specifically the series configuration, to model a relevant problem in the field of cerebral hemodynamic; it requires modeling the influence of CO_2 pressure ($PaCO_2$) on the system that regulates the cerebral blood flow. An increase of $PaCO_2$ in the body causes a state of hypercapnia, producing dilation of the blood vessels, deteriorating autoregulation, and increasing Cerebral Blood Flow Velocity (CBFV) [6]. The most widely used technique to measure the reactivity of the blood vessels of an individual to $PaCO_2$, consists in measuring the change produced in the CBFV by breathing a mixture of air and 5% $PaCO_2$ [7] to estimate the percent change in CBFV with respect to the change in $PaCO_2$ (the measurement of this ratio is known as the reactivity to $PaCO_2$). The blood flow velocity is measured with Transcranial Doppler Ultrasonography.

Since the CBFV variations depend also on the variations of Arterial Blood Pressure (ABP), the model is completed by measuring ABP in a noninvasive way on the middle finger with a Finapres instantaneous pressure gauge.

At present there are phenomenological models that approximately represent CBFV variation when there are changes in the levels of inspiration of $PaCO_2$ [8]. This model will be used as our *white box* model, to be part of the *gray box* method. Linear [7] and nonlinear nondeterministic models of the autoregulation phenomenon and the influence of $PaCO_2$ on this system have also been made [9-10]. In particular, the work that we have done [10] has shown that Support Vector Machines (SVM) represent an adequate paradigm for modeling (like a *black box*) the cerebral autoregulation system, under normal conditions and when $PaCO_2$ changes.

The hypothesis that we will prove in what follows has to do with a *gray box* model using the model of Poulin et al. [8] as *white box* and an SVM as *black box*, and it will allow a better representation of the phenomenon, both under normal conditions and under conditions of aspiration of 5% $PaCO_2$.

2 Methods

2.1 Data Collection

Sixteen healthy subjects aged 31.8±8.5 years were studied in a temperature controlled laboratory. None of them had a history of hypertension, diabetes, migraine, epilepsy, or any other cardiovascular or neurologic disease. The study was approved by the Leicestershire Research Ethics Committee and informed consent was obtained in all cases.

The subjects were asked to refrain from ingesting alcohol or caffeinated products in the 12 hours preceding the study. Measurements were made in the supine position. CBFV was recorded in the middle cerebral artery with transcranial Doppler (Scimed QVL-120) using a 2 MHz transducer. ABP was measured noninvasively using arterial volume clamping of the digital artery (Finapres 2300 Ohmeda). An infrared capnograph (Datex Normocap 200) with a face mask was used to measure end-tidal CO_2 ($EtCO_2$). The face mask was kept in place for the duration of the complete study including the $PaCO_2$ reactivity test with a mixture of 5% $PaCO_2$ in air administered

with a Douglas bag and elephant tubing connected to the face mask through a one-way valve.

Baseline values of CBFV, ABP and $EtCO_2$ were recorded for an initial period of 5 min with subjects breathing normal air, after all variables were stable for at least 15 min. This was followed by a 5 min. recording with each subject breathing a mixture of 5% $PaCO_2$ in air.

2.2 Pre-processing

All the signals were collected and saved on a digital audio tape using an 8-channel recording instrument (*Sony PC108M*), and they were then transferred to a microcomputer in real time. The fast Fourier transform was used to extract the maximum frequency of the CBFV signal, with a 5-ms time window. The signals were digitized and sampled at 200 samples/s, and then processed through an 8^{th} order zero-phase Butterworth low-pass filter with a cut-off frequency of 20 Hz.

The beginning and end of each cardiac cycle were detected in the arterial pressure signal, and the mean values of ABP and CBFV were calculated for each heart beat. Spline interpolation, followed by re-sampling every 0.2 s produced time series with a uniform time base. The $EtCO_2$ signal was interpolated linearly between successive end-tidal values and was also re-sampled at 0.2 s intervals. For the purpose of implementing SVM models, the signals were sub-sampled at 0.6 s intervals, resulting in approximately 500 data points for each of the two different segments of data.

2.3 White Box Model

The differential equation proposed by Poulin [8], which represents our *white box* model, is shown in Equation 1.

$$\frac{d(CBFV(t))}{dt} = \frac{1}{\tau}\left[g \bullet u(t - T_d) + CBFV^* - CBFV(t)\right] \qquad (1)$$

where the input $u(t\text{-}T_d)=[EtCO_2(t\text{-}T_d)\text{-}EtCO_2^*]$, and $EtCO_2^*$ is the control period. The three parameters g, τ, $CBFV^*$, are obtained using the least squares technique. In the case of constant T_d, a *grid* search is used that minimizes the sum of the squares of the other parameters. This equation is solved using separable variables to obtain the *CBFV(n)* in discrete times *n* which correspond to each heart beat.

2.4 Black Box Model

The adopted SVM algorithm was the ν-SVM, introduced by Vapnik in 1995 [11]. It is based on the statistical theory of learning, which introduced regression as the fitting of a tube of radius ε to the data. The decision boundary for determining the radius of the tube is given by a small subset of training examples called Support Vectors.

Assuming that $\vec{x}$ represents the input data vector, the output value $f(\vec{x})$ is given by the SVM regression using a weight vector $\vec{w}$, according to equation 2.

$$f(\vec{x}) = (\vec{w} \bullet \vec{x}) + b, \quad \vec{w}, \vec{x} \in \mathbf{R}^N, b \in \mathbf{R}, \tag{2}$$

where b is a constant obtained from $\vec{w}$.

The variation of the ν-SVM introduced by Schölkopf et al. [12] consists in adding ε to the minimization problem, weighted by a variable ν that adjusts the contribution of ε between 0 and 1.

$$\text{minimize} \quad \theta(\vec{w}, \xi) = \frac{1}{2}\left\|\vec{w}\right\| + C\left(l\nu\varepsilon + \sum_{i=1}^{l} \xi_i \right) \tag{3}$$

In equation 3, l represents the total dimension of the data (number of cases), C is a model parameter determining the trade-off between the complexity of the model, expressed by $\vec{w}$, and the points that remain outside the tube. Slack variables ξ depend on the distance of the data points from the regression line. We used the ε-insensitive loss function.

The solution of this minimization problem for obtaining the weight vectors $\vec{w}$ is found by the standard optimization procedure for a problem with inequality restrictions when applying the conditions of Kuhn-Tuker to the dual problem. The main advantage of introducing parameter $\nu \in [0\ 1]$ is to make it possible to control the error fraction and the number (or fraction) of Support Vectors with only one normalized parameter.

To solve a nonlinear regression problem it is sufficient to substitute the inner product between two independent original variables $\vec{x}_i \bullet \vec{x}_j$ (Eq. 2) by a kernel function gaussian radial base function (RBF), given by equation 4:

$$k(\vec{x}_i, \vec{x}_j) = \exp(-\left\|\vec{x}_i - \vec{x}_j\right\|^2 / (2\sigma^2)) \tag{4}$$

The implementation of the *black box* model for the case of a differential equation must correspond to a dynamic model in time. We have chosen a model of the AutoRegressive with Exogenous input (ARX) type that can consider one (ABP) or two (ABP and $EtCO_2$) inputs to model a parameter ($\theta =$ τ, g or *CBFV**) of Poulin's differential equation [8]. The *black box* model presents two options, as shown in equations 5 and 6.

$$\hat{\theta}(t) = f(\theta(t-1),..., \theta(t-n_\theta), p(t),..., p(t-n_p)) \tag{5}$$

$$\hat{\theta}(t) = f(\theta(t-1),..., \theta(t-n_\theta), p(t),..., p(t-n_p),..., c(t),..., c(t-n_c)) \tag{6}$$

where $p(t)=ABP(t)$, $c(t)=EtCO_2(t)$ and $\theta(t)$ is one of the parameters τ, g or *CBFV**.

Function $f()$ can be a linear function when Eq. 2 is used, or a nonlinear one when using the kernel RBF function shown in Eq. 4.

2.5 Gray Box Model

The *gray box* model is implemented using Poulin's differential equation as *white box*, for which some of the parameters τ, g or *CVFB** are estimated by means of an SVM as *black box*, as shown in Figure 1. When Eq. 5 is used, the dotted line from *c(t)* does not exist, and in the case of Eq. 6 the SVM has two inputs (*p(t)* and *c(t)*).

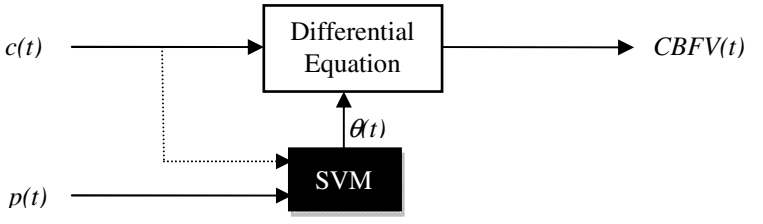

Fig. 1. *Gray box* model; the dotted line indicates if the *black box* model is univariate or multivariate

3 Results

Good quality recordings were obtained for all subjects with both the baseline and the 5% $PaCO_2$ test. Representative fluctuations in ABP, $EtCO_2$ and CBFV are shown in Figure 2 for one subject (#13) for the baseline and 5% $PaCO_2$ data segments.

To train and evaluate the different proposed models two states are chosen, one during the baseline period and the other during the aspiration of 5% de $PaCO_2$ in air. In each state one half of each period is chosen to train and the other half to evaluate. Since the output variables correspond to signals, we chose Pearson's correlation (r) between the real and the estimated CBFV signals, as an index to evaluate the precision of the models.

To estimate each of the three parameters, τ, g and *CBFV**, each of them is isolated from the solution of Eq. 1, and then the SVM models are trained and evaluated using one these parameters as output signals. The best results are obtained by modifying the *CBFV** parameter. The models are applied to each of the 16 subjects.

The results were calculated for the baseline and changes to 5% $PaCO_2$ conditions. In the baseline, for the univariate model of the SVM, only the nonlinear case is calculated. For multivariate 5% $PaCO_2$ changes only the nonlinear case is calculated (the linear cases that were not calculated are not significant under these conditions). The results of the average correlations for the 16 subjects are shown in Table 1.

Figure 3 shows the reactivity curves to $EtCO_2$ for the linear and nonlinear models under the 5% $PaCO_2$ change condition. It is also important to obtain the reactivity indices for these models, which are calculated as the ratio of the change between the CBFV values and the $EtCO_2$ changes. The average values for linear univariate 5% $PaCO_2$ is 4.8 (mm Hg/%), and 4.4 (mm Hg/%) for the nonlinear model.

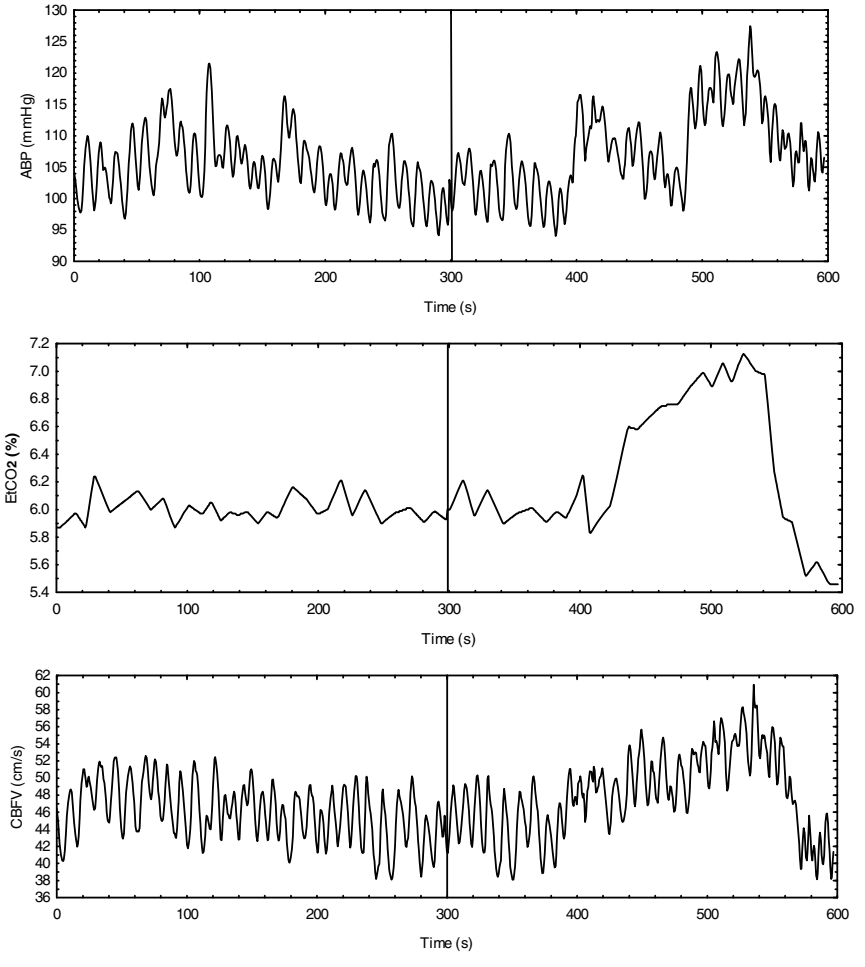

Fig. 2. Representative time-series of ABP, $EtCO_2$ and CBFV showing spontaneous fluctuations during baseline (left) and breathing 5% $PaCO_2$ in air (right)

Table 1. Average correlations for the baseline and 5% $PaCO_2$ change conditions for univariate and multivariate SVM models

Baseline		
Model	Training	Test
SVM Univariate		
Linear	-	-
Nonlinear	0.967	0.769#
SVM Multivariate		
Linear	0.968	0.727*
Nonlinear	0.967	0.801#*

Changes 5% CO_2		
Model	Training	Test
SVM Univariate		
Linear	0.989	0.948†
Nonlinear	0.986	***0.962***†‡
SVM Multivariate		
Linear	-	-
Nonlinear	0.987	0.951‡

When the Wilcoxon test was applied to establish if the differences were significant ($p<0.05$), the following values were obtained for p: #0.026, *0.002, †0.501, ‡0.535.

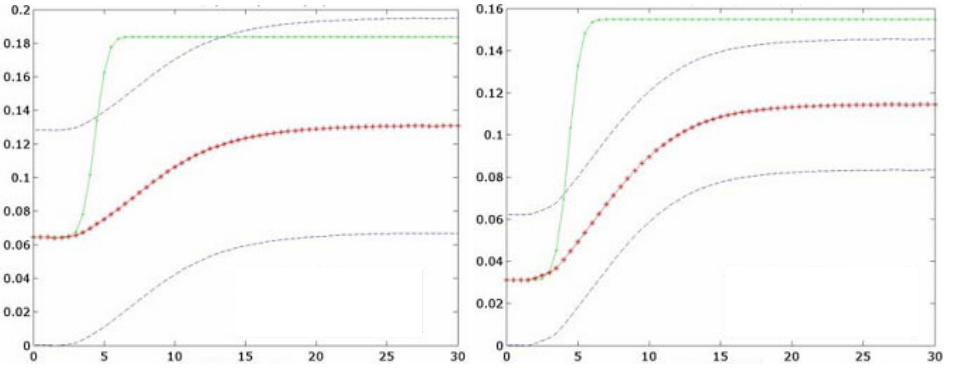

Fig. 3. Normalized CBFV time response (asterisks line) and its standard deviations (dashed line) to a change of $EtCO_2$ (dotted line). For the nonlinear multivariate model during baseline (left) and response for the nonlinear univariate model for breathing 5% $PaCO_2$ in air (right).

4 Discussion and Conclusions

When the selection of parameters of the differential equation to estimate them with the SVM is examined, the advantages of a *gray box* model can already be seen, because of the three parameters, the one that can have the largest variation is precisely the baseline of the CVBF*. It is this parameter the one that represents the output of the basic model of the Cerebral Blood Flow Autoregulation phenomenon, when the ABP is the input signal [6,13]. It is also interesting to note that it is the univariate model of the SVM (ABP→CBFV*) the one that achieves the best results, and there are no significant differences with the multivariate model of the SVM (ABP, EtCO2→CBFV*). This can be explained when it is considered that Poulin's equation [8] already considers the contribution of $EtCO_2$ on the variation of CBFV, so including it in the SVM is redundant.

The average reactivity curves shown in Fig. 3 as well as the calculated reactivity indices coincide with the values of normal subjects like those studied.

The excellent results obtained with the correlation index show that they are significantly better than the *white box* (*r= 0.805*) as well as the *black box* models presented in [10], (nonlinear: *r=0.707* for baseline and *r=0.909* for 5% $PaCO_2$), which were obtained with the same set of data.

The results shown in this *gray box* application with SVM to a problem of cerebral hemodynamic bring up the potential of the method in terms of precision as well as of the valuable contribution that can be obtained from the description of the phenomenon.

In future work it will be important to evaluate the contribution of the SVMs compared to Artificial Neural Networks, and the application to another field of science and engineering.

Acknowledgments. This work was supported by DICYT/USACH Project 061119CP.

References

1. Psichogios, D., Ungar, L.: A Hybrid Neural Networks First Principles Approach to Process Modeling. AIChE J. 38, 1499–1511 (1992)
2. Thompson, M., Kramer, M.: Modeling Chemical Processes Using Prior Knowledge and Neural Networks. AIChE J. 40, 1328–1340 (1994)
3. Anderson, J., McAvoy, T., Hao, O.: Use of Hybrid Models in Wastewater Systems. Ind. Eng. Chem. Res. 39, 94–1704 (2000)
4. Thibault, J., Acuña, G., Pérez-Correa, R., Jorquera, R., Molin, P., Agosin, E.: A hybrid representation approach for modelling complex dynamic bioprocesses. Bioprocess Engineering 22(6), 547–556 (2000)
5. Acuña, G., Cubillos, F., Thibault, J., Latrille, E.: Comparison of Methods for Training Grey- box Neural Models. Computers. Chem. Engng. 23, S561–S564 (1999)
6. Panerai, R.B.: Assesment of cerebral pressure autoregulation in humans - a review of measurement methods. Physiological Measurement 9, 305–338 (1998)
7. Simpson, D., Panerai, R., Evans, D., Garnham, J., Naylor, A., Bell, P.: Estimating normal and pathological flow velocity to step changes in end-tidal PCO2. Medical & Biological Engineering & Computing 38, 535–539 (2000)
8. Poulin, M., Liang, P., Robbins, P.: Dynamics of the cerebral blood flow response to step changes in end-tidal PCO2 and PO2 in humans. Journal of Applied Physiology 81, 1084–1095 (1996)
9. Mitsis, G., Poulin, M., Robbins, P., Marmarelis, V.: Nonlinear modeling of the dynamic effects of arterial pressure and CO2 variations on cerebral blood flow in healthy humans. IEEE Transactions on Biomedical Engineering 51, 1932–1943 (2004)
10. Chacón, M., Araya, C., Panerai, R.B.: Non-linear multivariate modeling of cerebral hemodynamics with autoregressive Support Vector Machines. Medical Engineering & Physics 33(2), 180–187 (2011)
11. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
12. Schölkopf, B., Smola, A., Williamson, R., Bartlett, P.: New support vector algorithms. Neural Computation 2, 1083–1121 (1998)
13. Tiecks, F., Lam, A., Aaslid, R., Newell, D.: Comparison of static and dynamic cerebral autoregulation measurements. Stroke 26, 1014–1019 (1995)