Improved Performance of Data-Adaptive Regression Framework Based on Multivariate Adaptive Regression Splines for Electrocardiographic Imaging

Amaël Mombereau[1], Yesim Serinagaoglu Dogrusoz[2], Ayoub El Ghebouli[1], Rémi Dubois[1], Laura Bear[1]

[1]Univ. Bordeaux, INSERM, CRCTB, U 1045, IHU Liryc, F-33000 Bordeaux, France
[2]Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey

Abstract

Noninvasive electrocardiographic imaging (ECGI) reconstructs cardiac electrical activity from body surface ECG. This study aims to 1) compare two ECGI methods, the traditional Tikhonov regularization with a nonparametric regression framework based on multivariate adaptative regression splines (MARS) using in-vivo experimental data, and 2) optimize how we use these data for training the MARS method. A regression model was trained with 184 body surface potentials and corresponding 239 unipolar electrograms (EGM) recorded from four anesthetized closed-chest pigs during sinus rhythm and pacing from various locations. A comparative analysis of four different training set compositions was performed on using a different number of beats. Performance of the methods was assessed by comparing estimated and true EGM and Activation Times (AT).

The MARS-based method had better fidelity to the original EGMs compared to Tikhonov regularization when the training data available was sufficient, in number of beats and number of stimulus sites. The best compromise between performance and computing time was to train using 5 beats for each pacing site.

1. Introduction

The goal of electrocardiographic imaging (ECGI) is to reconstruct cardiac electrical sources non-invasively from body surface potentials (BSP) [1]. This approach is used clinically to aid identification of ablation sites [2][3]. ECGI involves solving the inverse problem of electrocardiography (ECG). However, this problem is ill-posed meaning regularization is required [4-7], the most commonly used approach being Tikhonov regularization [8]. Despite significant progress, the outcomes remain less than optimal [9-11].

A novel nonparametric data-driven regression framework based on multivariate adaptive regression splines (MARS) has been proposed for ECGI [12]. Unlike traditional regularization techniques, data-driven inversion bypasses the need for an explicit forward model. Instead, it uses the data itself to uncover the relationship between input and output variables through a training process. The approach is structured into three key phases. Initially functions linking BSPs to individual EGMs are derived from a training dataset. Then, these functions are employed to map the measured BSPs onto the corresponding EGMs. Lastly spatial smoothing us used to enhance the precision of the spatiotemporal patterns in the reconstructed inverse solutions.

The MARS approach has previously been evaluated with simulated and torso-tank data, showing improved performance compared to a traditional Tikhonov approach. This study aims to extend this work by evaluating the accuracy of the MARS approach on *in-vivo* data and to optimize the data used for training in terms of accuracy vs computation time.

2. Methods

2.1. Data Sets

Experimental data were obtained from anesthetized, closed-chest pigs (n=4, 30-40 kg) [13] Epicardial and torso potentials were recorded simultaneously using a "sock" (239 unipolar electrodes, Auckland Uniservices Ltd, New Zealand) and flexible strips attached to the body surface (184 electrodes, BioSemi, the Netherlands). For each pig, recordings were made during sinus rhythm, and pacing from left and right endo- , and epicardial sites. In total 50 records were obtained of between 5 to 34 beats. Upon completion, the heart was arrested and MRI performed. A homogeneous heart-torso model was segmented from MRI as well as BSP electrode locations. The heart was excised and perfusion-fixed. Epicardial electrode locations were captured with a multi-axis digitizing arm (FARO Technologies, FL). MRI markers placed on the “sock” and body surface strips were used for registration.

2.2. Inverse Problem Formulation

In this work, we used extracellular electrograms (EGM) as the bioelectric source model. The relationship

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2024.294

between EGM and BSPs is described by (1):

$$y(t) = Ax(t) + n(t) \tag{1}$$

Here, $x(t) \in \mathbb{R}^{N\times T}$ and $y(t) \in \mathbb{R}^{M\times T}$ are the EGM and BSP vectors at time t, where M, N and T are the number of ECG electrodes, epicardial nodes and number of time samples, respectively. $A \in \mathbb{R}^{M\times N}$ is the forward operator, and $n(t) \in \mathbb{R}^{M\times T}$ represents the measurement noise. The forward matrix A was calculated by solving the forward ECG problem using the boundary element method. For the inverse problem, zero-order Tikhonov regularization was employed as a benchmark for comparison with the MARS method. The regularization parameter was computed using the L-curve method.

2.3. Multivariate Adaptive Regression Splines (MARS)

The MARS method has previously been described in [12]. MARS is a non-parametric regression technique designed to model complex relationships between a dependent variable and multiple predictors [14]. It achieves this by using spline functions to divide the data into different regions and fit local models in each interval. The central idea of MARS is to construct the model through a series of basis functions formed around knots, which act as breakpoints in the predictor space. These basis functions are defined as pairs of reflected splines and can capture both non-linearities and interactions between variables. For a given predictor x and knot τ, the basic form of these splines is expressed as:

$$(x - \tau)_+ = \max(0, x - \tau) \tag{2}$$

where τ represents the knot location and x is a scalar input variable. The algorithm starts by defining a training set consisting of BSP and EGM measurement pairs. Then, MARS automatically generates the knot points. The MARS model is constructed in two main steps. First, in the forward phase, the algorithm iteratively adds pairs of basis functions that most reduce the residual sum of squares (RSS), a measure of the error between the observed data and the model's predictions. This process begins with a simple model containing only the intercept and progressively adds complexity by incorporating splines that improve the fit. The RSS is defined as follows:

$$RSS = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \tag{3}$$

where y_i is the observed response, $f(x_i)$ is the predicted value from the MARS model, and n is the number of observations (BSP-EGM measurement vector pairs).

Once the forward phase concludes, the model may be overly complex. To address this, MARS applies a backward pruning process, where non-essential basis functions are removed to avoid overfitting. This pruning is guided by the Generalized Cross-Validation (GCV) criterion, which balances the goodness of fit against the complexity of the model. The GCV is calculated as:

$$GCV = \frac{RSS}{\left(1 - \frac{C(M)}{n}\right)^2} \tag{4}$$

where $C(M)$ is a complexity cost function. The goal is to select a model that minimizes GCV, thus preventing overfitting while maintaining predictive accuracy. An important feature of MARS is its ability to model interactions between predictors. This is achieved by introducing products of basis functions, such as $B(x, \tau)$. This approach allows MARS to capture higher-order interactions between variables. MARS automatically determines the location of the knots by searching for points that best reduce the error. During the backward pruning phase, the algorithm eliminates functions that contribute minimally to improving the model’s accuracy, thus simplifying the overall structure.

We implemented our methods with the help of “Earth: Multivariate Adaptive Regression Splines R package” and its standalone C version, which can be used in MATLAB [15]. After constructing the regression model, it is applied to estimate EGMs from a new set of BSP measurements. However, the spatial maps of the reconstructed EGMs may display irregularities due to insufficient spatial correlation between the mapping functions. To address this, spatial smoothing is performed to reduce fluctuations in potential values between neighboring points on the epicardial surface. We utilized a geometrically smooth spline-based data-fitting procedure (MARS-based) on the reconstructed EGM to achieve this smoothing, which can be applied generally without the need for customization to specific beats or datasets.

2.4. Training Data

The model was trained on each pig, leaving one pacing site out (unique pacing location) for the test dataset, and using the rest for training. A comparative analysis of four different training set compositions was performed on the MARS-based method using: 1) a single beat for each pacing site (single), 2) five beats (five), 3) all available beats (all-beats) and 4) a signal-averaged beat (average).

2.5. Data analysis

Data analysis was performed on all beats for each recording. Correspondence between measured and reconstructed EGMs was quantified using the relative error (RE) and Pearson’s correlation coefficient (CC).

Activation times (AT) were estimated from EGMs using the minimum derivative approach. The correspondence of measured and reconstructed AT was also compared by evaluating CC (AT-CC). For each metric, the significance of differences between the means of the inverse methods were determined using a one-way ANOVA or Friedman test.

3. Results

3.1. Comparison with Tikhonov and MARS-based methods

The proposed MARS-based method was compared with Tikhonov method. Results are presented in Table.1, summarized in terms of median and interquartile range [IQR] for all metrics. In general, the estimation accuracy of MARS was significantly higher ($p<0.0001$) but varied more than Tikhonov method (larger IQR). However, when the database was too small, MARS does not perform as well as Tikhonov (e.g. for Pig 2).

Table 1: Summary of results obtained for MARS-based and Tikhonov regularization method in terms of CC, RE and AT-CC. Values are given as median [IQR], best in green and worst in red.

Pig	site nb	Method	CC	RE	AT-CC
1	11	MARS	0,87 [0,32]	0,51 [0,45]	0,88 [0,26]
		Tik	0,56 [0,03]	0,82 [0,06]	0,77 [0,19]
2	5	MARS	0,34 [0,18]	1,60 [2,38]	0,47 [0,36]
		Tik	0,62 [0,12]	0,81 [0,13]	0,76 [0,22]
3	13	MARS	0,84 [0,10]	0,56 [0,12]	0,82 [0,19]
		Tik	0,66 [0,12]	0,76 [0,07]	0,75 [0,15]
4	21	MARS	0,71 [0,25]	0,72 [0,34]	0,70 [0,17]
		Tik	0,50 [0,16]	0,86 [0,04]	0,65 [0,07]
all	50	MARS	0,74 [0,25]	0,69 [0,54]	0,73 [0,25]
		Tik	0,55 [0,12]	0,51 [0,45]	0,67 [0,14]

Fig.1 presents two examples of true and reconstructed AT during (A) endocardial RV apex and (B) epicardial RV base pacing. The AT maps reconstructed shown in Fig.1A reveal, in the best case, that the MARS model produced almost identical patterns to the true map (CC=0.98). Whereas in Fig.1B, the worst case, there is no similarity at all (CC=0.18). This poor performance is likely due to the lack of training data, having only trained with 4 of 5 available pacing sites (pig 2).
Tikhonov generally did not perform as well as MARS, creating line of block artefact where it did not exist. However results were more stable, always capturing the general propagation pattern. In terms of EGM reconstruction, Tikhonov regularization produced an over-smoothed signal and could not determine sharp changes in the potential values, while MARS-based method tended to produce EGMs that were similar to the ground truth, but noisier.

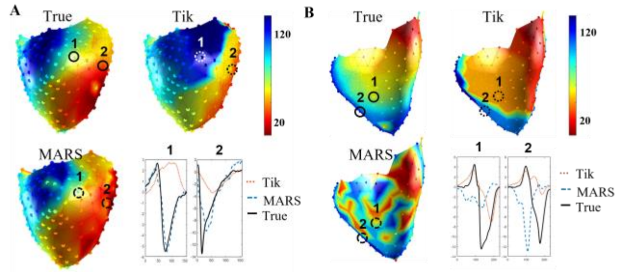
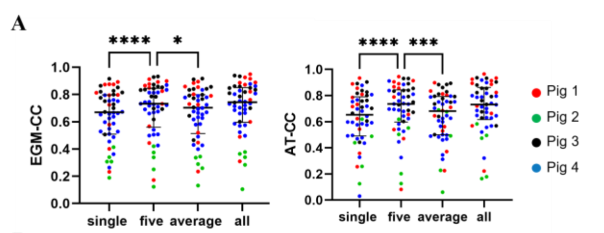


Figure 1: True and reconstructed activation time maps and EGMs, in two different localization, for all methods. (A) the best and (B) the worst case for MARS-based method.

3.2. Optimizing MARS-Based method Training: balancing accuracy and efficiency

The CC of EGM and AT reconstructed with the MARS-based method increased with an increasing number of training beats (Fig 2A), as did the computation time (Fig 2B). Using a single-averaged beat was slightly better than using single beat (though no significantly), while keeping computation low, thirty-fold lower than training with all beats. The best compromise between performance and computing time would be to train over 5 beats, being as accurate as using all beats and was significantly better than the single or averaged beat training, with a computing time that remained relatively low.



B

Number of beat	Single	five	average	all
Computation time (s)	120 [18]	567 [137]	122 [32]	3340 [1060]

*Figure 2: Comparative analysis of four different training set for MARS based method, using: single beat, five beats, all available beats and a signal-averaged beat. (A) CC of EGM and AT for 4 different experiments (pig). (B) Computing time for each training set. With **** $p\leq0.0001$, ***$p\leq0.001$, *$p\leq0.05$*

4. Discussion

This study provides novel insights into the performance and applicability of MARS-based framework for ECGI. The MARS-based method, when supplied with sufficient training data, clearly outperforms the traditional Tikhonov regularization approach in terms of EGMs with higher fidelity. Even without quantitative detection, a notable advantage of MARS seems to be its ability to avoid the line of block artefact, leading to smoother and more physiologically accurate AT reconstructions.

However, the study also reveals that MARS's performance is highly dependent on the quantity and diversity of training data. When the dataset is limited, Tikhonov regularization may offer more consistent results. An option to improve the stability may be to couple *in-vivo* and *in-silico* datasets for training. The optimization of MARS, through the use of multiple or averaged beats during training, also shows promise in enhancing both accuracy and computational efficiency.

In this study, the MARS-based model was trained using data from multiple pacing sites within the same pig, demonstrating its ability to accurately reconstruct electrograms (EGM) and activation times (AT) within that specific subject. However, the ultimate goal is to develop a model that can predict EGM and AT not just within the same subject but across different subjects. To enhance the model's generalizability, future work will focus on registering data from the different pigs to allow a “leave one pig-out” approach for the test dataset. This approach would significantly broaden the model's clinical applicability.

5. Conclusion

We demonstrated that the Multivariate Adaptive Regression Splines (MARS)-based framework provides a more accurate reconstruction of ECGI when sufficient training data is available, compared to the traditional Tikhonov regularization method. The MARS-based method not only improves the fidelity of reconstructed electrograms (EGMs) but also avoids the introduction of line of block artifacts that can occur in Tikhonov regularization. The results suggest that MARS-based techniques, when optimized the training data set using 5 beats for each pacing site, offer a promising alternative for enhancing the precision of ECGI.

Acknowledgments

This research was financially supported by the National Research Agency (ANR-10-IAHU-04) and (ANR-22-CE17-0023), and the PHC BOSPHORE “INVERSE” grant (47978NM), and Scientific and Technological Research Council of Türkiye (TÜBİTAK-221N175).

References

[1] R. M. Gulrajani, The forward and inverse problems of electrocardiography , *IEEE Eng. Med. Biol. Mag.*, vol. 17, nº 5, p. 84-101, 122, oct. 1998, doi: 10.1109/51.715491.

[2] J. Zhang, K. A. Desouza, P. S. Cuculich, D. H. Cooper, J. Chen, et Y. Rudy, Continuous ECGI mapping of spontaneous VT initiation, continuation, and termination with antitachycardia pacing , *Heart Rhythm*, vol. 10, nº 8, p. 1-2, 2012.

[3] P. S. Cuculich *et al.*, Noninvasive characterization of epicardial activation in humans with diverse atrial fibrillation patterns , *Circ.*, vol. 122, nº 14, p. 1364-72, 2010.

[4] L. K. Bawua, C. Miaskowski, X. Hu, G. W. Rodway, et M. M. Pelter, A review of the literature on the accuracy, strengths, and limitations of visual, thoracic impedance, and electrocardiographic methods used to measure respiratory rate in hospitalized patients , *Ann. Noninvasive Electrocardiol.*, vol. 26, nº 5, p. e12885, sept. 2021, doi: 10.1111/anec.12885.

[5] A. Rahimi, J. Xu, et L. Wang, -Norm Regularization in Volumetric Imaging of Cardiac Current Sources , *Comput. Math. Methods Med.*, vol. 2013, p. 1-10, 2013, doi: 10.1155/2013/276478.

[6] B. Erem, J. Coll-Font, R. Martinez Orellana, P. St’ovicek, et D. H. Brooks, Using Transmural Regularization and Dynamic Modeling for Noninvasive Cardiac Potential Imaging of Endocardial Pacing With Imprecise Thoracic Geometry , *IEEE Trans. Med. Imaging*, vol. 33, nº 3, p. 726-738, mars 2014, doi: 10.1109/TMI.2013.2295220.

[7] T. Erenler et Y. Serinagaoglu Dogrusoz, ML and MAP estimation of parameters for the Kalman filter and smoother applied to electrocardiographic imaging , *Med. Biol. Eng. Comput.*, vol. 57, nº 10, p. 2093-2113, oct. 2019, doi: 10.1007/s11517-019-02018-6.

[8] A. N. Tikhonov et V. Y. Arsenin, *Solutions of ill-posed problems*. Washington, D.C.: John Wiley & Sons, New York: V. H. Winston & Sons, 1977.

[9] M. J. M. Cluitmans *et al.*, In Vivo Validation of Electrocardiographic Imaging , *JACC Clin. Electrophysiol.*, vol. 3, nº 3, p. 232-242, mars 2017, doi: 10.1016/j.jacep.2016.11.012.

[10] K.-W. Chen, L. Bear, et C.-W. Lin, Solving Inverse Electrocardiographic Mapping Using Machine Learning and Deep Learning Frameworks , *Sensors*, vol. 22, nº 6, p. 2331, mars 2022, doi: 10.3390/s22062331.

[11] A. S. Revishvili *et al.*, Validation of the mapping accuracy of a novel non-invasive epicardial and endocardial electrophysiology system , *Europace*, vol. 17, nº 8, p. 1282-1288, août 2015, doi: 10.1093/europace/euu339.

[12] O. Onak, T. Erenler, et Y. Serinagaoglu, A Novel Data-Adaptive Regression Framework Based on Multivariate Adaptive Regression Splines for Electrocardiographic Imaging , *IEEE Trans. Biomed. Eng.*, vol. 69, nº 2, p. 963-974, févr. 2022, doi: 10.1109/TBME.2021.3110767.

[13] L. R. Bear *et al.*, Forward Problem of Electrocardiography: Is It Solved? , *Circ. Arrhythm. Electrophysiol.*, vol. 8, nº 3, p. 677-684, juin 2015, doi: 10.1161/CIRCEP.114.001573.

[14] J. H. Friedman et C. B. Roosen, An introduction to multivariate adaptive regression splines , *Stat. Methods Med. Res.*, vol. 4, nº 3, p. 197-217, sept. 1995, doi: 10.1177/096228029500400303.

[15] S. Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani., earth: Multivariate Adaptive Regression Splines . p. 5.3.3, 19 avril 2007. doi: 10.32614/CRAN.package.earth.

Address for correspondence:

Amaël Mombereau
Av. du Haut Lévêque, 33600 Pessac
Amael.mombereau@ihu-liryc.fr