



# Méthodologie Pour Comparer Différentes Méthodes D'extraction De Biomarqueurs Sans Méthode De Référence. Application À La Segmentation Du Ventricule Gauche En Irm Cardiaque Pour Estimer La Fraction D'éjection.

Frédérique Frouin, Mireille Garreau, Irène Buvat, Chistopher Casta, Constantin Constantinidès, Jean Cousty, Alexandre Cochet, Stéphanie Jehan-Besson, Christophe Tilmant, Murielle Lefort, et al.

## ► To cite this version:

Frédérique Frouin, Mireille Garreau, Irène Buvat, Chistopher Casta, Constantin Constantinidès, et al.. Méthodologie Pour Comparer Différentes Méthodes D'extraction De Biomarqueurs Sans Méthode De Référence. Application À La Segmentation Du Ventricule Gauche En Irm Cardiaque Pour Estimer La Fraction D'éjection.. RITS 2011 (Colloque National Recherche en Imagerie et Technologies pour la Santé), Apr 2011, Rennes (CHU), France. IRBM numéro spécial RITS 2011, 2011.

**HAL Id: hal-00908741**

**<https://hal.archives-ouvertes.fr/hal-00908741>**

Submitted on 3 Jan 2017

teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Méthodologie pour comparer différentes méthodes d'extraction de biomarqueurs sans méthode de référence. Application à la segmentation du ventricule gauche en IRM cardiaque pour estimer la fraction d'éjection

Frédérique Frouin<sup>1</sup>, Mireille Garreau<sup>2</sup>, Irène Buvat<sup>3</sup>, Christopher Casta<sup>4</sup>, Constantin Constantinidès<sup>1,5</sup>, Jean Cousty<sup>6</sup>, Alexandre Cochet<sup>7</sup>, Stéphanie Jehan-Besson<sup>8</sup>, Christophe Tilmant<sup>9</sup>, Muriel Lefort<sup>1</sup>, Laurent Najman<sup>6</sup>, Laurent Sarry<sup>10</sup>, Patrick Clarysse<sup>4</sup>, Alain de Cesare<sup>1</sup>, Alain Lalande<sup>7</sup>

<sup>1</sup>Inserm U678 ; UPMC UMR\_S 678, Laboratoire d'Imagerie Fonctionnelle, Paris

<sup>2</sup>Inserm UMR 642 ; Université Rennes 1, Laboratoire du Traitement du Signal et de l'Image, Rennes

<sup>3</sup>CNRS UMR 8165 ; Université Paris-Sud 11, IMNC, Orsay

<sup>4</sup>Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon, Villeurbanne

<sup>5</sup>ESME-Sudria, Laboratoire PRIAM, Ivry sur Seine

<sup>6</sup>Université Paris Est, UMR 8049, Laboratoire d'Informatique Gaspard-Monge, A3SI, ESIEE, Paris

<sup>7</sup>CNRS UMR 5158 ; Université de Bourgogne, Laboratoire d'Electronique, d'Informatique et d'Image, Dijon

<sup>8</sup>CNRS UMR 6158, LIMOS, Aubière

<sup>9</sup>CNRS UMR 6602 ; Université Blaise Pascal, LASMEA, Clermont-Ferrand

<sup>10</sup>Université d'Auvergne, ERIM EA 3295, Clermont-Ferrand

**Résumé** – Une méthodologie est proposée pour comparer plusieurs méthodes d'estimation d'un paramètre d'intérêt clinique sans connaître de gold standard. Elle est appliquée à un problème de comparaison de la fraction d'éjection du ventricule gauche en IRM cardiaque obtenue par sept approches de segmentation différentes et appliquée à 30 examens issus de la base de données constituée pour le challenge Miccai 2009 de segmentation du ventricule gauche. La comparaison des méthodes montre une supériorité du tracé manuel, un bon comportement des méthodes de segmentation semi-automatiques et un plus mauvais comportement des méthodes plus largement automatisées. La méthodologie de comparaison proposée qui ne fait aucun a priori sur les méthodes retrouve ainsi l'ordre attendu. Cette approche méthodologique est donc pertinente et doit permettre une classification non supervisée de différentes méthodes d'estimation de biomarqueurs.

## Mots Clefs

Modélisation statistique, Evaluation non supervisée, Segmentation, IRM cardiaque, Fraction d'éjection

## I. INTRODUCTION

La comparaison d'algorithmes de segmentation sur des bases d'images communes est une question cruciale et d'actualité. La plupart des approches font l'hypothèse d'une méthode de référence, celle-ci étant le plus souvent apportée par la délimitation réalisée par un « expert » ou un compromis obtenu à partir de différents experts.

Nous proposons un cadre méthodologique pour effectuer ces comparaisons en l'absence de référence à partir de l'approche « RWT » de régression linéaire en l'absence de

vérité [1, 2]. Cette approche est modifiée en définissant une nouvelle figure de mérite [3] et en proposant une classification des différentes méthodes sur une analyse de rang faite après une procédure de bootstrap sur l'ensemble des échantillons. L'ensemble est mis en œuvre pour comparer différentes méthodes de segmentation du ventricule gauche en IRM en vue d'estimer la fraction d'éjection. Cette étude a été réalisée dans le cadre de l'action MedIEval (Medical Image segmentation Evaluation) du GdR STIC-Santé.

## II. THEORIE

Considérons de façon générale  $P$  jeux de données, (indiqués par  $p$ ,  $p$  variant de 1 à  $P$ ) à partir desquels  $M$  méthodes (indiquées par  $m$ ,  $m$  variant de 1 à  $M$ ) fournissent l'estimation d'un biomarqueur  $\theta_{pm}$  ; la valeur réelle  $\Theta_p$  de ce biomarqueur est inconnue.

Le modèle « RWT » [1, 2] repose sur les trois hypothèses suivantes :

i) la distribution de probabilité du biomarqueur  $\Theta$  sur la population des  $P$  jeux de données suit une loi à support fini (en l'occurrence une loi bêta) ;

ii) chaque méthode  $m$  conduit à une estimée  $\theta_{pm}$  de  $\Theta_p$ , telle qu'il existe une relation linéaire entre  $\theta_{pm}$  et  $\Theta_p$ , à un bruit près (1) :

$$\theta_{pm} = a_m \Theta_p + b_m + \varepsilon_{pm}, \quad (1)$$

$\varepsilon_{pm}$  suivant une loi normale centrée, d'écart-type  $\sigma_m$  ;

iii) les erreurs commises par chaque méthode sont indépendantes.

Sous ces hypothèses, il est possible d'exprimer la probabilité conjointe  $\Pr\{\theta_{pm}, \Theta_p\}$  par l'équation (2):

$$\Pr(\theta_{pm} / \{a_m, b_m, \sigma_m\}, \Theta_p) = \prod_{m=1}^M \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2\sigma_m^2} \cdot (\theta_{pm} - a_m \Theta_p - b_m)^2\right) \quad (2)$$

puis de spécifier  $\Pr(\theta_{pm} / \{a_m, b_m, \sigma_m\})$ . En considérant l'ensemble des  $P$  jeux de données, le logarithme de la vraisemblance s'écrit finalement en fonction de  $a_m$ ,  $b_m$  et  $\sigma_m$  et des paramètres de la loi de probabilité de  $\Theta$  [1]. La maximisation de cette expression permet d'estimer pour chaque méthode les paramètres  $a_m$ ,  $b_m$  et  $\sigma_m$  et les paramètres  $\mu$  et  $\nu$  de la loi bêta.

La première figure de mérite, proposée dans [1, 2], vaut  $F_1 = \sigma_m / a_m$ . Nous l'avons testée vis-à-vis d'une seconde figure de mérite  $F_2$ , définie dans [3] comme l'espérance de l'écart quadratique entre la valeur du paramètre et la valeur estimée:  $E(\Theta - a_m \Theta - b_m - \varepsilon_m)^2$ . En supposant les erreurs indépendantes de  $\Theta$  et en considérant l'expression des moments d'ordre 1 et 2 d'une loi bêta, il est possible de montrer que  $F_2$  s'écrit:

$$F_2 = (a_m - 1)^2 \frac{\mu(\mu + 1)}{(\mu + \nu)(\mu + \nu + 1)} + 2 \cdot (a_m - 1) b_m \frac{\mu}{\mu + \nu} + \sigma_m^2$$

(3)

Pour estimer de façon robuste les figures de mérite et considérant le cas où l'échantillon de jeux de données est de faible effectif, nous avons réalisé une procédure de bootstrap [4] sur les biomarqueurs estimés à partir de chaque méthode en réalisant 1000 tirages différents de  $P$  échantillons  $\{\theta_{p_1}, \theta_{p_2}, \dots, \theta_{p_p}\}$ ,  $\theta_{p_i}$  étant l'ensemble des  $M$  valeurs  $\theta_{m, p_i}$  associé au jeu de données  $p_i$ . A partir des figures de mérite  $F_1$  et  $F_2$  calculées pour chaque tirage, un test de Kruskal-Wallis sur les rangs des  $M$  méthodes a été effectué et une comparaison des méthodes a été établie après correction par l'approche de Bonferroni pour les tests multiples.

### III. MATERIELS ET METHODES

Cette méthodologie a été mise en œuvre en considérant comme paramètre  $\Theta$  la fraction d'éjection estimée à partir de différentes méthodes ( $M=7$ ) de segmentation du ventricule gauche. Nous avons utilisé la base de données fournie par le challenge Miccai en 2009 [5], en retenant les 30 études ( $P=30$ ) des phases de validation et de test « on line », incluant des cas sains et différents cas pathologiques. Chaque examen est une acquisition IRM (séquence SSFP) comportant une dizaine de coupes petit-axe du cœur, acquises sur 20 phases du cycle cardiaque. Pour l'estimation du paramètre de fraction d'éjection, les temps de diastole et de systole et les coupes sur lesquelles la segmentation devait se faire ont été imposés aux différents participants pour que la variabilité dans les résultats obtenus ne soit pas liée à ces choix.

Cinq laboratoires ont participé à la segmentation des données, au total sept jeux de contours ont été obtenus permettant sept estimations indépendantes de la fraction d'éjection. Deux des méthodes (M1 et M7) sont des segmentations manuelles réalisées par des experts de deux laboratoires différents. Les cinq autres méthodes de segmentation automatisée ont été décrites dans [6, 7, 8]. Trois des méthodes (M2, M5 et M6) ont été utilisées de façon semi-automatique avec des possibilités d'intervention et de reprise de paramètres de la part de l'opérateur. Deux d'entre elles ont été reprises avec un paramétrage différent et en n'autorisant qu'une interaction très limitée de l'opérateur, conduisant aux méthodes de segmentation appelées M3 et M4.

### IV. RÉSULTATS

La figure 1 montre le résultat de la procédure « RWT » sur les sept méthodes d'estimation de la fraction d'éjection, les figures de mérite  $F_1$  et  $F_2$  associées sont indiquées en Table 1.

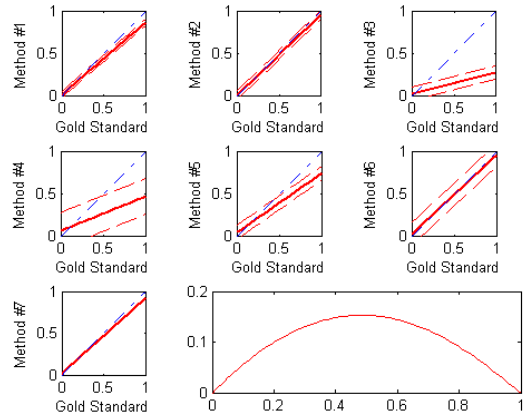


Fig.1: Les sept premiers quadrants montrent l'estimation de la régression linéaire (en rouge) des sept méthodes vis-à-vis du gold standard (pointillés bleus), les pointillés rouges représentant l'écart-type lié à la régression; le dernier quadrant représente l'estimation de la densité de probabilité de la fraction d'éjection sur l'échantillon de 30 jeux de données.

Méthode	$F_1$	$F_2$
M1	0,0368	0,0057
M2	0,0454	0,0032
M3	0,2541	0,1543
M4	0,4548	0,1162
M5	0,1000	0,0225
M6	0,1058	0,0182
M7	0,0025	0,0013

TABLE 1: Figures de mérite obtenues pour les sept méthodes, correspondant à l'estimation montrée sur la Fig.1.

La méthode de comparaison des sept méthodes a été appliquée à  $F_1$  et  $F_2$ , elle est illustrée pour  $F_2$ . La figure 2 donne la répartition de la figure de mérite  $F_2$  associée à chacune des sept méthodes suite à la procédure de bootstrap. La figure 3 illustre la comparaison des 7

méthodes après analyse de Kruskal-Wallis sur les rangs obtenus suite aux 1000 tirages, après correction de Bonferroni pour les tests multiples, en prenant la limite sur l'erreur de première espèce égale à 5%.

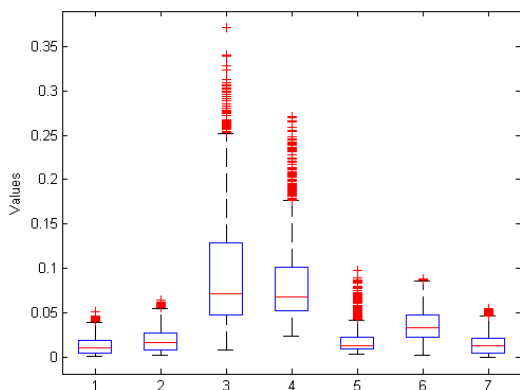


Fig.2 : « Boxplot » pour chacune des sept méthodes (abscisses) de la figure de mérite  $F_2$  (ordonnées) suite à la procédure de bootstrap (1000 tirages) : médiane (trait rouge), intervalle entre le 1<sup>er</sup> et 3<sup>ème</sup> quartile (rectangle bleu), valeurs distantes d'au maximum 1,5 fois la distance interquartile (pointillés noirs), points atypiques (+ rouge)

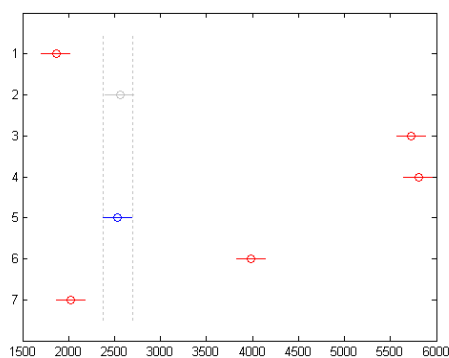


Fig.3 Comparaison des rangs moyens x1000 (abscisses) obtenus pour chacune des sept méthodes (ordonnées) à partir de la figure de mérite  $F_2$ . Ainsi, la méthode M5 (en bleu) est dans le même groupe que la méthode M2 (située dans l'espace délimité par les traits pointillés, traduisant les intervalles de confiance).

Pour la figure de mérite  $F_1$ , six groupes de méthodes sont mis en évidence distinguant des plus performantes aux moins performantes : M7, M1, M2, M5-M6, M3, M4. Pour la figure de mérite  $F_2$ , quatre groupes de méthodes sont estimés, donnant des plus performantes aux moins performantes: M1-M7, M2-M5, M6, M3-M4.

## V. DISCUSSION – CONCLUSION

Les deux figures de mérite produisent des résultats sensiblement équivalents : il ressort un très bon comportement des méthodes M1 et M7 qui correspondent aux segmentations manuelles, les méthodes les moins performantes étant les méthodes M3 et M4, c'est-à-dire celles faisant le moins appel à l'opérateur. Or ce classement correspond à l'idée attendue et au jugement

visuel qui peut être établi sur la qualité des contours extraits. Ainsi, la méthodologie proposée qui ne fait aucun *a priori* autre que les trois hypothèses indiquées en section II (et donc aucune hypothèse sur l'ordre des méthodes) semble pertinente pour comparer différentes approches d'estimation de la fraction d'éjection. Cependant, dans ce résumé, les méthodes n'ont pas été identifiées explicitement, pour ne pas aboutir à une classification définitive reposant sur le seul critère de la fraction d'éjection et sans validation supplémentaire de l'approche « RWT ».

De plus, au sein du groupe MedIEval, nous cherchons à compléter cette approche par une méthodologie permettant de classer les méthodes plus finement, patient par patient. Une approche permettant d'intégrer l'ensemble des segmentations réalisées pour définir une forme robuste moyenne [9] reposant sur une formulation dérivée de la méthodologie STAPLE [10] est en cours de développement.

## REMERCIEMENTS

Les auteurs remercient le GdR 2647 Stic-Santé pour le soutien qu'il apporte à l'action MedIEval.

## REFERENCES

- [1] J.W. Hoppin, M.A. Kupinski, G.A. Kastis, E. Clarkson, H.H. Barrett, Objective comparison of quantitative imaging modalities without the use of a gold standard, *IEEE Trans Med Imaging*, Vol. 21, pp. 441-449, 2002.
- [2] M.A. Kupinski, J.W. Hoppin, J. Krasnow, S. Dahlberg, J.A. Leppo, M.A. King, E. Clarkson, H.H. Barrett, Comparing cardiac ejection fraction estimation algorithms without a gold standard, *Acad Radiol*, Vol. 3, pp 329-337, 2006.
- [3] M. Soret, J. Alaoui, P.M. Koulibaly, J. Darcourt, I. Buvat, Accuracy of partial volume effect correction in clinical molecular imaging of dopamine transporter using SPECT, *Nuclear Instruments and Methods in Physics Research A*, Vol. 571, pp.173-176, 2007.
- [4] B. Efron, R.J. Tibshirani *An introduction to the Bootstrap*, Chapman et Hall, 1993.
- [5] Cardiac MR Left Ventricle Segmentation Challenge [http://smial.sri.utoronto.ca/LV\\_Challenge/Home.html](http://smial.sri.utoronto.ca/LV_Challenge/Home.html)
- [6] J. Schaerer, C. Casta, J. Pousin, P. Clarysse, A dynamic elastic model for segmentation and tracking of the heart in MR image sequences, *Medical Image Analysis*, Vol. 14, pp. 738-749, 2010.
- [7] C. Constantinidès, Y. Chenoune, E. Mousseaux, F. Frouin, E. Rouillot, Automated Heart Localization for the Segmentation of the Ventricular Cavities on Cine Magnetic Resonance Images, *Computing in Cardiology*, Vol. 37, pp. 911-914, 2010.
- [8] J. Cousty, L. Najman, M. Couprie, S. Clément-Guinaudeau, T. Goissen, J. Garot, Segmentation of 4D cardiac MRI: automated method based on spatio-temporal watershed cuts, *Image and Vision Computing*, Vol. 28, pp. 1229-1243, 2010.
- [9] S. Jehan-Besson, C Tilmant Estimation variationnelle de formes moyennes, Journées MedIEval, mai 2010
- [10] S.K. Warfield, K.H. Zou, W.M. Wells, Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, *IEEE Trans Med Imaging*, Vol. 23, pp. 903-921, 2004.