



ISTIA
École
d'ingénieurs

Master Systèmes Dynamiques et Signaux

Mémoire

Objets connectés et suivi à domicile de patients traités pour un cancer

Auteur :

M. Antoine JAMIN

Encadrant :

Dr. Jean-Baptiste FASQUEL

Jury :

Pr. L. HARDOUIN

Pr. A. HUMEAU-HEURTIER

Dr. JB. FASQUEL

Dr. P. RICHARD

Version du
28 juin 2017

Remerciements

Tout d'abord je tiens à remercier Jean-Baptiste FASQUEL, mon tuteur qui m'a initié au monde de la recherche et qui m'a permis de réaliser mon projet dans les meilleures conditions ainsi que de préparer mon avenir en me conseillant sur les différentes propositions de thèse qui s'offraient à moi.

Je tiens également à remercier Samir HENNI, docteur au CHU d'Angers, pour sa disponibilité, ses précieux conseils et explications ainsi que pour m'avoir recommandé lorsque j'ai postulé à une thèse.

Je tiens également à remercier Anne HUMEAU-HEURTIER, pour ses précieux conseils qui m'ont permis de réaliser mon projet dans les meilleures conditions ainsi que pour m'avoir conseillé et aidé à définir mon projet professionnel.

Je tiens également à remercier l'ISTIA, le LARIS et leurs personnels pour m'avoir permis de faire ce master et ce stage importants pour mon avenir professionnel.

Pour finir, je tiens à remercier mes parents et ma sœur qui m'ont soutenu tout au long de l'année.

Table des matières

Introduction	1
1 Le projet	3
1.1 Les objets connectés dans le domaine de la santé	3
1.2 Le protocole DyVaa	4
1.3 La méthodologie envisagée	7
1.4 Environnement de travail	7
2 Analyse de séries temporelles	9
2.1 Descripteurs Statistiques	9
2.2 Dynamic Time Warping	12
2.3 Entropie	17
2.4 Analyse de la première phase du traitement	22
2.5 Résultats	26
Conclusion	29
Annexe	35

Table des figures

1.1	Déroulement du traitement.	5
1.2	Signaux représentatifs de l'évolution de la PAM.	6
1.3	Signaux représentatifs de l'évolution de la PAS et de la PAD.	6
2.1	Descripteurs statistiques d'ordre 1 et 2.	10
2.2	Descripteurs statistiques d'ordre 3 et 4.	11
2.3	Vecteurs de caractérisation 3-D.	11
2.4	Comparaison entre la distance euclidienne et la DTW.	13
2.5	Construction du dendrogramme par clustering hiérarchique ascendant. . . .	14
2.6	Matrice de distance.	14
2.7	Dendrogramme des signaux de la PAM, utilisant la DTW.	15
2.8	PLA appliquée à nos signaux.	15
2.9	Dendrogramme des signaux de la PAM, utilisant la DTW à la PLA.	16
2.10	RCMSE : Signal en grain grossier.	18
2.11	RCMSE des signaux de références de la Figure 1.2.	20
2.12	RCMSE et clusters.	20
2.13	Indice de complexité de tous les patients.	21
2.14	Entropie de l'échantillon croisée de nos patients.	22
2.15	Profils théoriques de l'évolution de la PAM au cours du traitement.	22
2.16	Taux de dépassement de la phase 1 en fonction du coefficient directeur a . .	24
2.17	Groupe 1 de l'analyse de la première phase du traitement.	25
2.18	Groupe 2 de l'analyse de la première phase du traitement.	25
2.19	Profils des patient qui nous posent problème.	26

Liste des acronymes

VEGF	<i>Vascular Endothelial Growth Factor</i>
DyVaa	<i>Dysfonction Vasculaire induite par les AntiAngiogénique</i>
CHU	<i>Centre Hospitalier Universitaire</i>
IoT	<i>Internet of Things ou Internet des Objets</i>
MSE	<i>MultiScale Entropy ou Entropie multi-échelle</i>
RCMSE	<i>Refined Composite MultiScale Entropy</i>
PAS	<i>Pression Artérielle Systolique</i>
PAD	<i>Pression Artérielle Diastolique</i>
PAM	<i>Pression Artérielle Moyenne</i>
DTW	<i>Dynamic Time Warping</i>
PLA	<i>Piecewise Linear Approximation</i>
IC	<i>Indice de Complexité</i>

Introduction

Le cancer du rein est le douzième type de cancer le plus fréquent dans le monde. 338 000 nouveaux cas ont été détectés en 2012 [1]. Ce type de cancer peut être traité par une chimiothérapie de type VEGF¹ qui consiste à éliminer les capillaires de la tumeur. Cependant, ce type de traitement a comme effet secondaire de faire augmenter la pression artérielle. C'est pourquoi un traitement antihypertenseur peut être prescrit aux patients. Notre projet s'adresse aux personnes atteintes d'un cancer du rein et traitées par chimiothérapie de type VEGF. Tout au long du traitement, les patients sont suivis tous les jours par le biais d'un tensiomètre connecté. Les objets connectés facilitent le suivi de l'évolution de la pression artérielle chez les patients. Notre but dans cette étude est d'analyser les données émises par ces objets connectés pour les mettre en corrélation avec le diagnostic médical.

Les objets connectés concernent tous les dispositifs qui peuvent être identifiés, faire des mesures et accomplir des tâches tout en étant ouvert sur internet. Ils représentent une discipline émergente dans le monde car on peut estimer qu'en 2020 le nombre d'objets connectés en circulation s'élèvera entre 50 et 80 milliards [2]. Les objets connectés dans le domaine de la santé connaissent une forte expansion. Tous ces dispositifs collectent un nombre de données important, c'est pourquoi il est important de s'intéresser à l'analyse de celles-ci. Ces données sont apparentées aux séries temporelles dont il existe un nombre important de méthodes d'analyse en fonction des caractéristiques propres de chaque série. Dans ce rapport nous aborderons quelques unes de ces méthodes.

La première partie de ce rapport nous présentera le projet de recherche sur lequel nous avons travaillé durant ce stage. Nous verrons tout d'abord les objets connectés dans le domaine de la santé, puis nous présenterons le protocole DyVaa et pour finir nous parlerons de la méthodologie envisagée et de l'environnement de travail. La seconde partie de ce rapport détaillera toutes les méthodes d'analyse de séries temporelles avec lesquelles nous avons travaillé ainsi que les résultats obtenus pour chacune de ces méthodes.

1. Vascular Endothelial Growth Factor

Chapitre 1

Le projet

Ce chapitre a pour but de faire un rappel sur le travail bibliographique réalisé au premier semestre. Nous résumerons donc dans un premier temps le contexte des objets connectés dans le domaine de la santé. Dans un second temps nous rappellerons le protocole DyVaa sur lequel nous travaillons. Nous détaillerons par la suite la méthodologie utilisée pour mener à bien notre étude. Pour finir, nous parlerons de l'environnement de travail de ce stage de recherche.

1.1 Les objets connectés dans le domaine de la santé

Comme le soulignent les récentes études [3, 4], le domaine des objets connectés évolue continuellement depuis une vingtaine d'année. Il constitue un enjeu majeur futur dans le domaine des nouvelles technologies.

L'internet des objets (IoT) est défini par le fait que des objets puissent être identifiés, faire des mesures, accomplir des tâches tout en étant reliés à un réseau ouvert sur internet. Les objets connectés sont donc accessibles où l'on veut à n'importe quelle échelle fixée. Ils servent également à collecter des informations ce qui laisse apparaître une nouvelle discipline émergente, le "Big Data".

Il est important de faire une différence entre capteurs connectés, qui ont une architecture simple dans le seul but de collecter des données, et objets connectés dont l'architecture est plus complexe dans le but de collecter et traiter les données.

Comme le soulignent les récentes études [5, 6, 7], les objets connectés (IoT) dans le domaine de la santé sont en pleine expansion avec pour but principal de collecter un maxi-

mum de données dans le but d’accompagner et d’améliorer le diagnostique médical ainsi que d’accompagner le patient dans le traitement de sa maladie.

Le domaine le plus émergent à l’heure actuelle est pour l’accompagnement des personnes à travers des objets connectés grands-publics de type montres connectées pour suivre son activité cardiaque, balances connectées pour suivre sa masse corporelle et bien plus encore [8]. Le milieu des personnes âgées est un domaine emprunt à une évolution importante ces dernières années [9].

Dans le domaine de l’oncologie, l’IoT est encore aux prémices d’un avenir certain [10]. Nous pouvons noter l’émergence de projets assurant le suivi de patients en cours de traitement cancéreux ainsi que pour d’autres types de maladie [11, 12, 13, 14].

1.2 Le protocole DyVaa

1.2.1 Présentation

Notre projet consiste à suivre le traitement de patients atteints de cancer du rein par le biais d’un tensiomètre connecté permettant de récupérer la pression artérielle plusieurs fois par jour de manière autonome. A l’heure actuelle le médecin reçoit ces données instantanément et le système émet une alerte (envoi d’un mail) à chaque fois que la tension dépasse un seuil fixé au préalable par le médecin. Les seuils fixés dans notre cas sont de 140 mmHg pour la pression artérielle systolique (PAS) et de 90 mmHg pour la pression artérielle diastolique (PAD). Ce dispositif de suivi n’est pour le moment pas optimum car les médecins se retrouvent très rapidement submergés par une quantité d’alerte trop importante.

Le traitement suivi est illustré par la Figure 1.1. Le traitement est appliqué lors de la découverte (C0) d’une tumeur cancéreuse (représentée en hachuré sur la figure) au niveau du rein. Ensuite elle est traitée par chimiothérapie (C1) de type VEGF qui tente de détruire les capillaires de la tumeur (représentés en rouge) pour empêcher son expansion (C1). Tout au long du traitement, le patient est suivi à distance par le tensiomètre connecté, il l’est également par le biais de consultations régulières (entre 3 et 6 visites pour un traitement de 3 mois (C2, C3, ...)).

Le traitement a comme effet de faire augmenter la pression artérielle chez le patient. Cette augmentation est signe que le traitement fonctionne bien, cependant une trop forte augmentation conduit à des risques importants non négligeables (AVC, Infarctus, ...). Ce

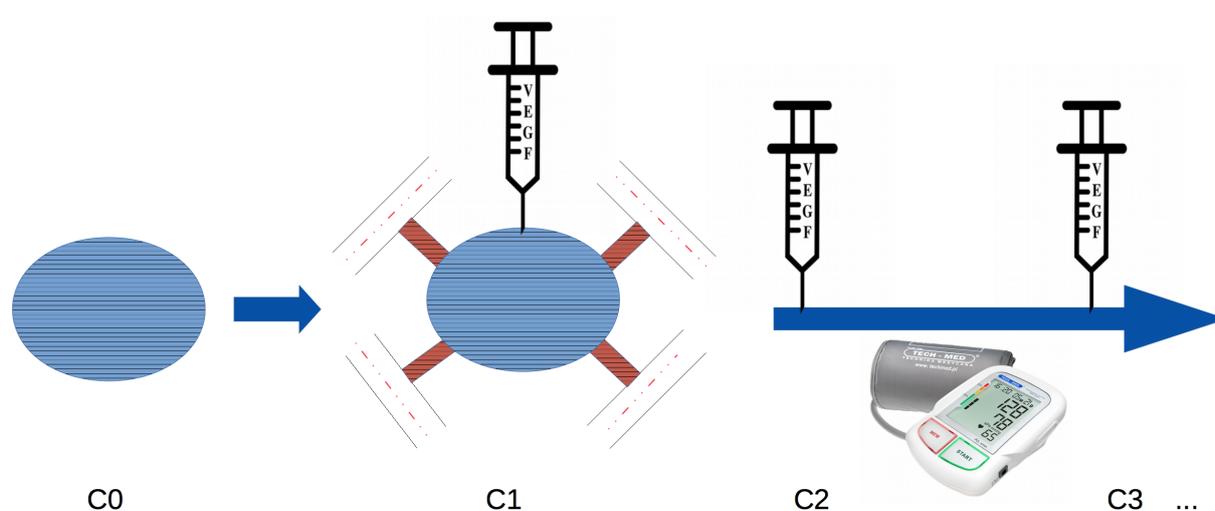


FIGURE 1.1 – Déroulement du traitement.

point nous montre tout l'intérêt de pouvoir affiner notre modèle de suivi de tension pour pouvoir mieux adapter le traitement tout en minimisant les risques secondaires. Pour lutter contre cette hausse importante de la pression artérielle, des antihypertenseurs peuvent être administrés dès le début du traitement (C1).

1.2.2 Les données du projet

Pour cette étude 28 patients, dont 8 suivis par télé-monitoring, effectuent une prise de leur tension artérielle suivant un processus de mesure commun. Ces acquisition nous permettent de récupérer la PAS et la PAD. Ces deux paramètres nous permettent de calculer la pression artérielle moyenne (PAM), définie par :

$$PAM = \frac{PAS + 2 \times PAD}{3}. \quad (1.1)$$

La Figure 1.2 nous montre trois courbes représentatives de l'évolution de la PAM des patients. Nous pouvons voir qu'il existe plusieurs profils de courbe. La Figure 1.3 nous montre les trois courbes représentatives de l'évolution de la PAS et de la PAD. Nous pouvons voir que le nombre de dépassements de seuil varie en fonction des patients.

1.2.3 L'objectif

L'objectif est de voir si la tension, mesurée quotidiennement, permet de distinguer des groupes ayant des propriétés particulières selon d'autres variables. A terme, l'objectif sera de prédire l'évolution de la pathologie, d'évaluer au plus tôt l'efficacité du traitement, et

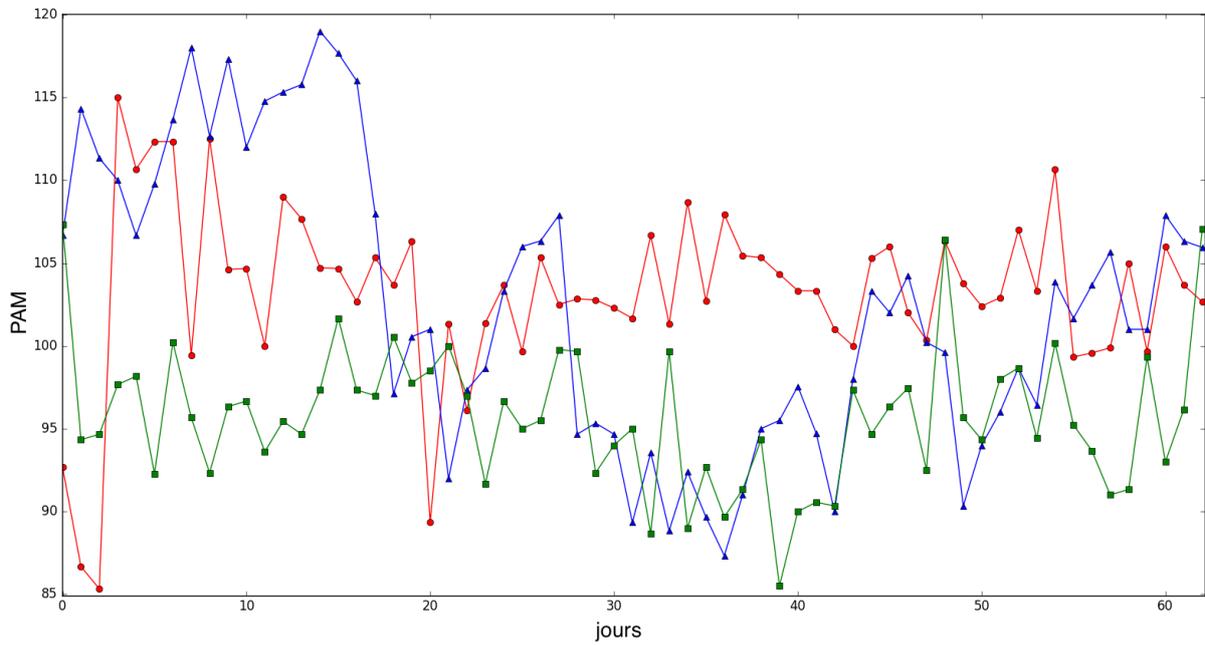
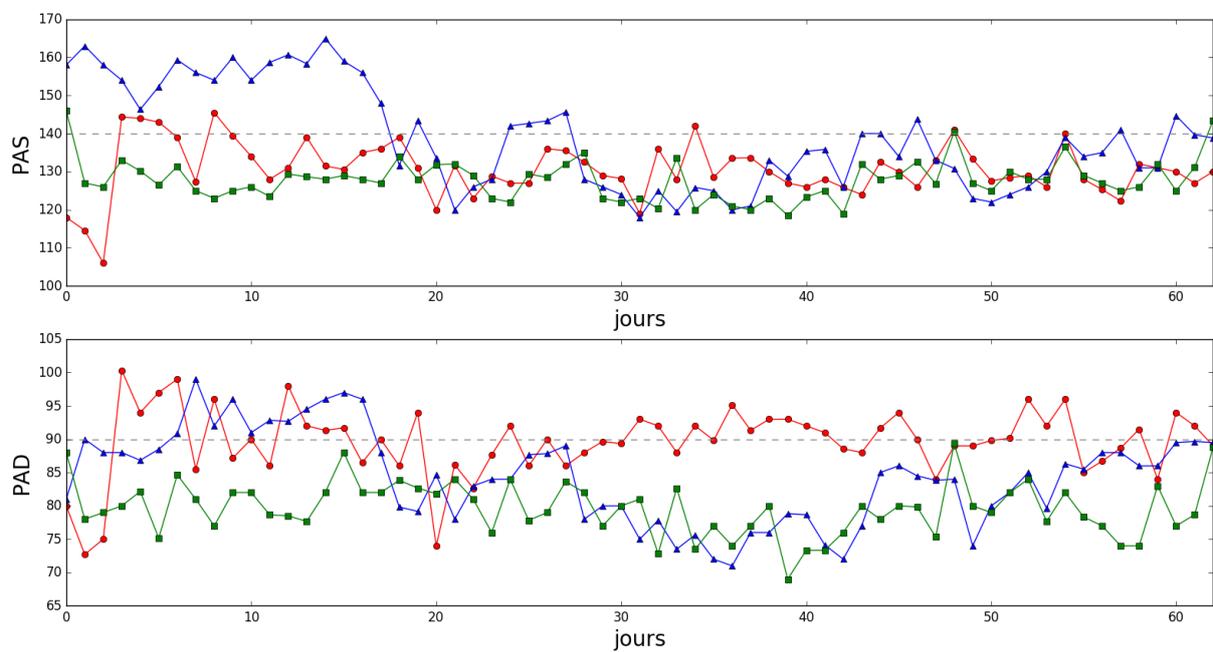


FIGURE 1.2 – Signaux représentatifs de l'évolution de la PAM.

FIGURE 1.3 – Signaux représentatifs de l'évolution de la PAS et de la PAD.
Les seuils sont mis en évidence par les lignes grises en pointillés

d'affiner la levée d'alerte. Ceci inclut également la possibilité d'aider les médecins lors de la mise en place de la procédure de suivie. Par exemple, il serait utile de pouvoir identifier les patients nécessitant un suivi quotidien prolongé, en fonction de l'évolution de la tension durant une période limitée en début de traitement.

1.3 La méthodologie envisagée

Comme le montre les récentes études [15, 16], il existe de nombreuses méthodes d'analyse et de clustering des séries temporelles. Nous avons décidé d'en étudier plusieurs :

- Descripteurs statistiques simples, avec, en particulier, les moments d'ordre 3 et 4.
- *Dynamic time warping* (DTW) : mesure de similarité/dissimilarité entre deux signaux en considérant le paramètre temporel comme non linéaire. Dans ce cas, nous considérerons une étude sur les données brutes et une étude après approximation linéaire par morceaux, a priori moins sensible aux bruits.
- Entropie : on s'intéressera à distinguer les signaux en fonction de leur irrégularité, à différentes échelles dans le cas de l'entropie multi-échelle.

Chaque méthode admet ses propres descripteurs. Nous avons privilégié une observation visuelle des groupes formés par ceux-ci. Ceci nous permet de nous affranchir des limitations intrinsèques des techniques automatiques de clustering telles que les *k-means*.

En complément de ces méthodes d'analyse de séries temporelles, nous étudierons également la fréquence des dépassements de seuil en début de traitement. Nous les comparerons avec les dépassements en fin de traitement. Il s'agit de voir si l'on peut trouver un lien entre l'évolution de la tension en début de traitement et celle de fin de traitement, avec l'idée de pouvoir décider de la stratégie de télé-monitoring, sur la base de premières mesures.

1.4 Environnement de travail

Ce projet de recherche est un projet de collaboration entre le pôle exploration vasculaire du CHU d'Angers et l'équipe de recherche ISISV du LARIS. Une partie de l'étude a été consacrée à la communication entre ces deux parties par le biais de compte-rendus et rapports réguliers ainsi que des réunions permettant de faire le point sur l'avancement du projet.

Toutes les méthodes de traitements des données ont été développées en Python à l'aide de nombreuses bibliothèques que nous avons adaptées à nos problèmes.

Chapitre 2

Analyse de séries temporelles

Dans ce chapitre nous allons présenter les différentes méthodes utilisées dans ce projet, nous présenterons également le résultat global de ces méthodes.

2.1 Descripteurs Statistiques

Dans le but de caractériser nos signaux par un ensemble de données, nous nous sommes intéressés aux descripteurs statistiques (DS) de chaque signal. Comme décrit dans [17, 18], nous avons décidé de prendre la moyenne (μ), décrite par l'équation 2.1, l'écart-type (σ), décrit par l'équation 2.2, l'asymétrie (*skewness*), décrite par l'équation 2.3 et l'aplatissement (*kurtosis*), décrit par l'équation 2.4. Ces deux derniers paramètres vont nous permettre d'avoir l'information sur la forme de la série temporelle. Ces quatre paramètres constituent un vecteur de caractérisation (*feature vector*) pour chaque signal. C'est sur ces vecteurs que nous allons pouvoir appliquer des algorithmes de clustering.

$$\mu = \frac{\sum_{t=1}^n y(t)}{n}, \quad (2.1)$$

$$\sigma = \sqrt{\frac{\sum_{t=1}^n (y(t) - \mu)^2}{n}}, \quad (2.2)$$

$$SKEW = \frac{\sum_{t=1}^n (y(t) - \mu)^3}{n\sigma^3}, \quad (2.3)$$

$$KURT = \frac{\sum_{t=1}^n (y(t) - \mu)^4}{n\sigma^4} - 3, \quad (2.4)$$

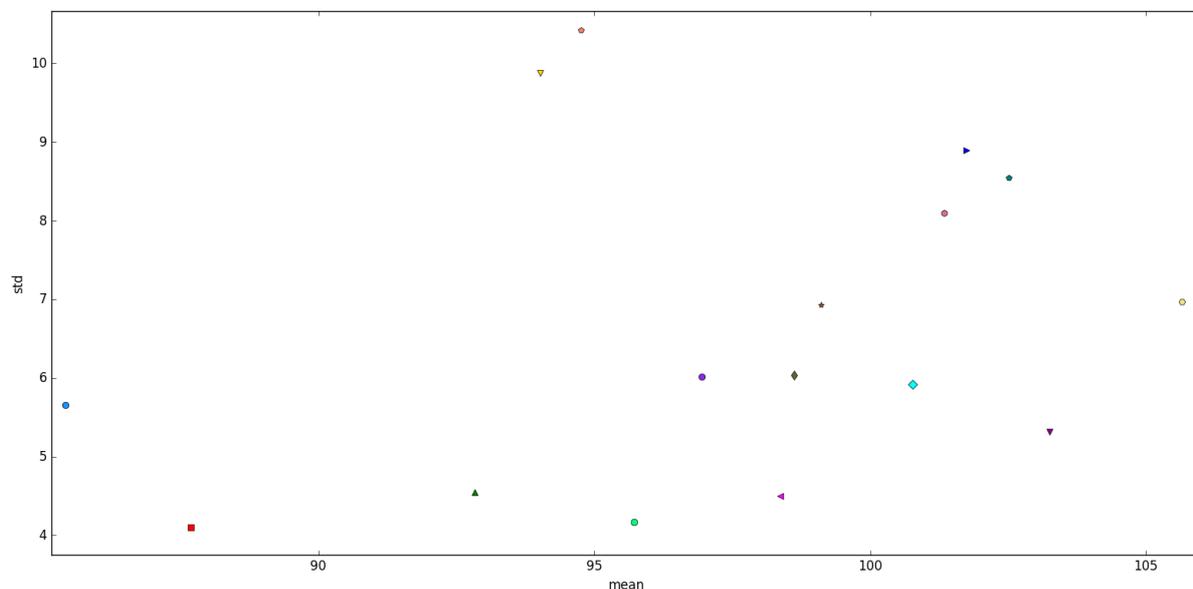


FIGURE 2.1 – Descripteurs statistiques d'ordre 1 et 2.

où $y(t)$ est le signal de taille n .

La Figure 2.1 montre les descripteurs statistique d'ordre 1 et 2 et la Figure 2.2 montre ceux d'ordre 3 et 4.

La valeur moyenne de la pression artérielle varie d'un individu à l'autre, c'est pourquoi dans un premier temps nous avons décidé de créer un vecteur de caractérisation 3-dimensions avec pour chaque individu l'écart-type, l'asymétrie et l'aplatissement de sa PAM. La Figure 2.3 montre la représentation de ces vecteurs de caractérisation.

Nous avons également, pour voir l'intérêt des 4 descripteurs statistiques (présentés ci-dessus), créé un vecteur de caractérisation 4-dimensions avec pour chaque individu l'asymétrie, l'aplatissement, l'écart-type et la moyenne de sa PAM.

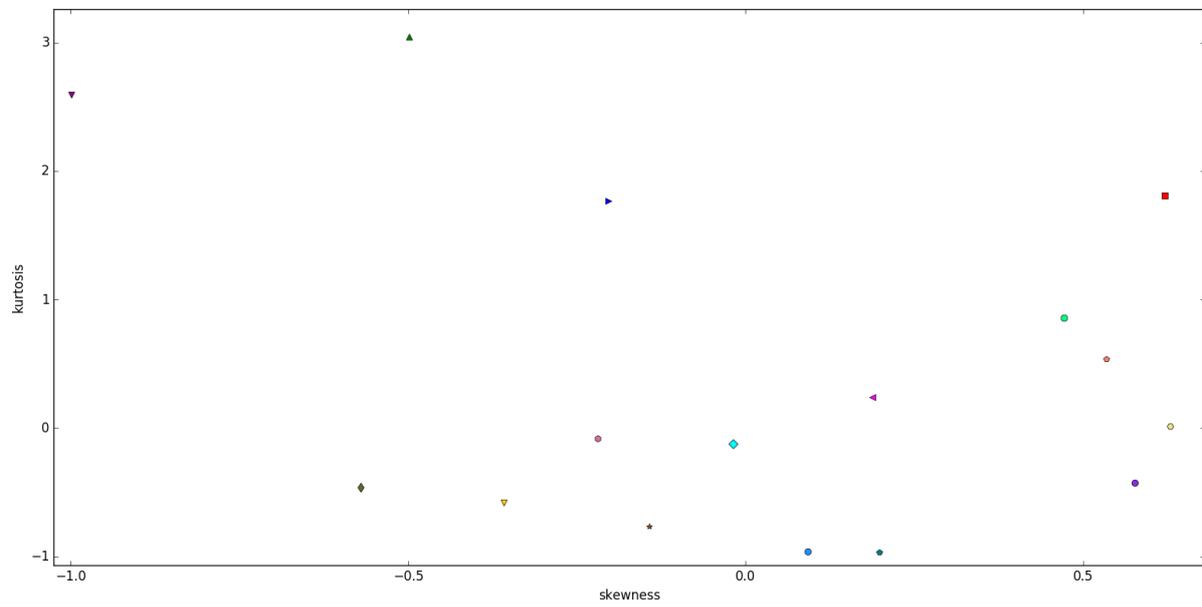


FIGURE 2.2 – Descripteurs statistiques d'ordre 3 et 4.

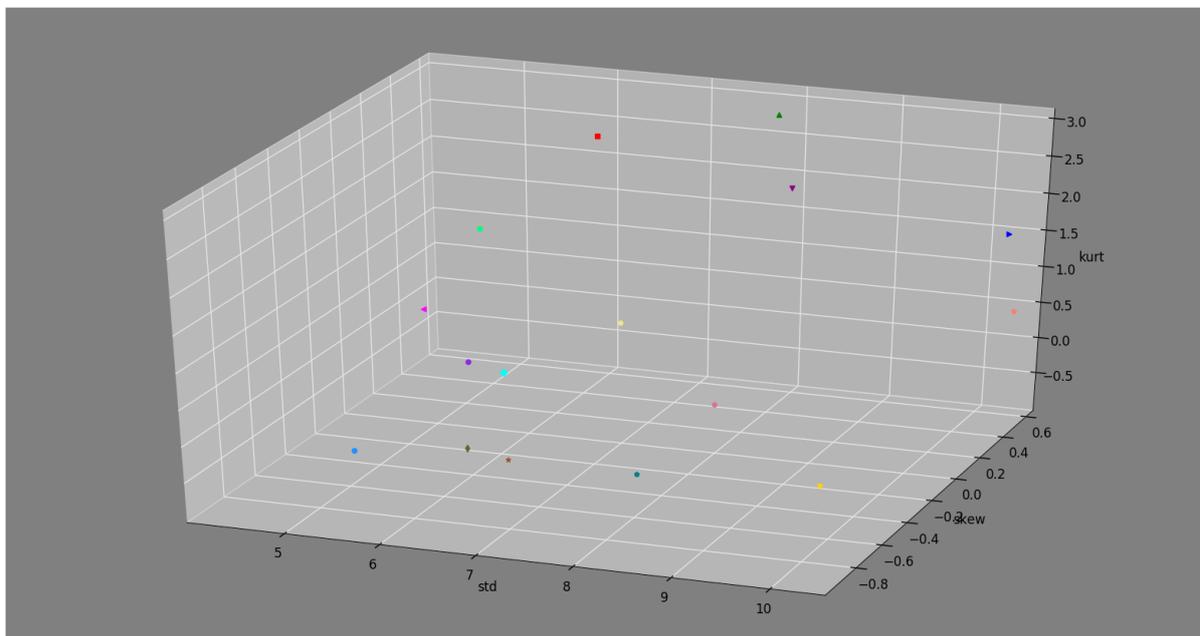


FIGURE 2.3 – Vecteurs de caractérisation 3-D.

2.2 Dynamic Time Warping

2.2.1 Méthode

L'algorithme du Dynamic Time Warping, comme indiqué dans [19, 20], permet de faire une mesure de distance entre deux signaux. Cette mesure considère le temps comme non linéaire et élastique ce qui permet d'obtenir un calcul de distance des signaux prenant en compte leurs décalages temporels.

Dans notre cas il est important de prendre en compte ce décalage temporel car la réaction au traitement n'est temporairement pas la même pour chaque patient. Par exemple, deux patients réagissant bien au traitement vont avoir une PAM qui augmente cependant, il peut y avoir un décalage de quelques jours entre ces deux patients ce qui montre l'intérêt d'utiliser une mesure de distance qui prend en compte ces décalages.

La valeur moyenne de la pression artérielle varie d'un individu à l'autre, c'est pourquoi nous devons normaliser nos signaux pour ne prendre en compte que les variations. Nous avons normalisé les signaux de la PAM en utilisant l'équation suivante :

$$x_n = \frac{x - \bar{x}}{\sigma_x}, \quad (2.5)$$

avec x le signal d'origine, \bar{x} la moyenne du signal d'origine et σ_x l'écart-type du signal d'origine.

La notion de prise en compte des décalages temporels est illustrée par la Figure 2.4 montrant la différence entre la distance euclidienne (ne prenant pas en compte le décalage) et la DTW.

2.2.2 Clustering hiérarchique

Le clustering hiérarchique consiste à construire une suite de clusters emboîtés les uns dans les autres [20]. Ces clusters seront donc mis en lien par le biais d'un arbre appelé également dendrogramme. Cette technique est très utile lorsqu'on ne connaît pas à l'avance le nombre de cluster que nous cherchons. La Figure 2.5 illustre la construction d'un arbre dans le cas d'un clustering hiérarchique ascendant. Pour construire ce dendrogramme, il nous faut choisir une mesure de distance (euclidienne, DTW, ...) et un critère d'agrégation (lien minimum, maximum ou moyen).

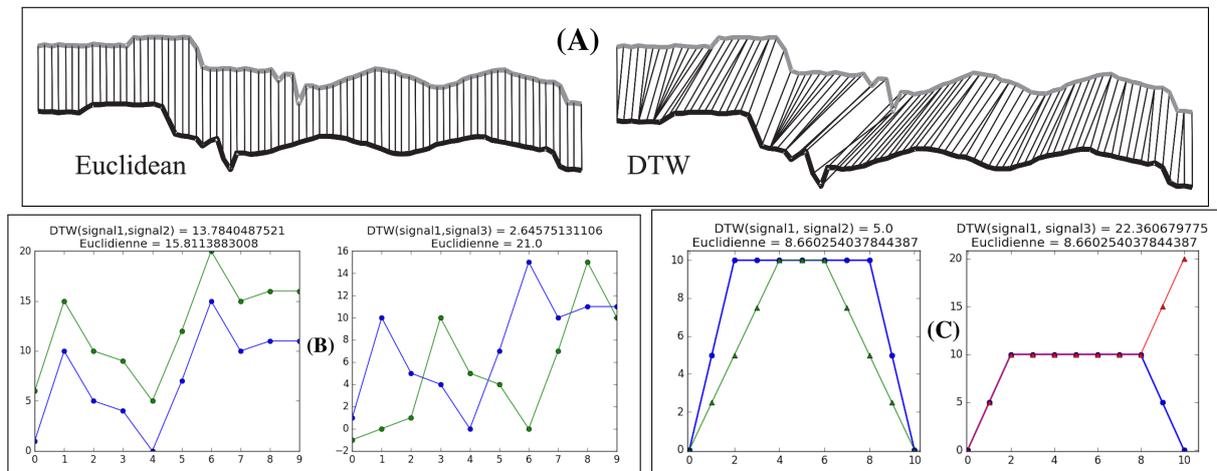


FIGURE 2.4 – Comparaison entre la distance euclidienne et la DTW.

(A) illustre la différence entre le calcul de la distance euclidienne et la DTW (figure extraite de [21]). (B) nous montre que les deux calculs de distance sont de même ordre de grandeur lorsque les signaux sont alignés et qu'ils sont différents lorsqu'ils sont déphasés.

La DTW prend donc en compte le déphasage dans le calcul. (C) nous montre que la DTW prend en compte l'allure des signaux dans le calcul de la distance.

A l'aide de la DTW, nous avons pu obtenir une matrice de distance illustrée par la Figure 2.6. Cette matrice représente la distance entre chaque signal de la PAM pour chaque patient.

Pour le clustering hiérarchique appliqué à nos données, nous avons choisi d'utiliser la DTW comme mesure de dissimilarité. Pour regrouper les patients à l'aide de la matrice de distance nous avons choisi le critère du lien moyen qui utilise la moyenne des distances entre les éléments de chaque classe pour effectuer les regroupements [20].

Le dendrogramme obtenu est représenté à l'aide de la Figure 2.7.

2.2.3 Approximation linéaire par morceaux

D'après [22, 23], l'approximation linéaire par morceaux – *Piecewise Linear Representation (PLA)* – consiste à représenter un signal par un ensemble de segments. Le nombre de segments varie en fonction du signal car il dépend uniquement d'une marge d'erreur fixée au préalable. Cette technique nous permet de filtrer le bruit (variations très court terme

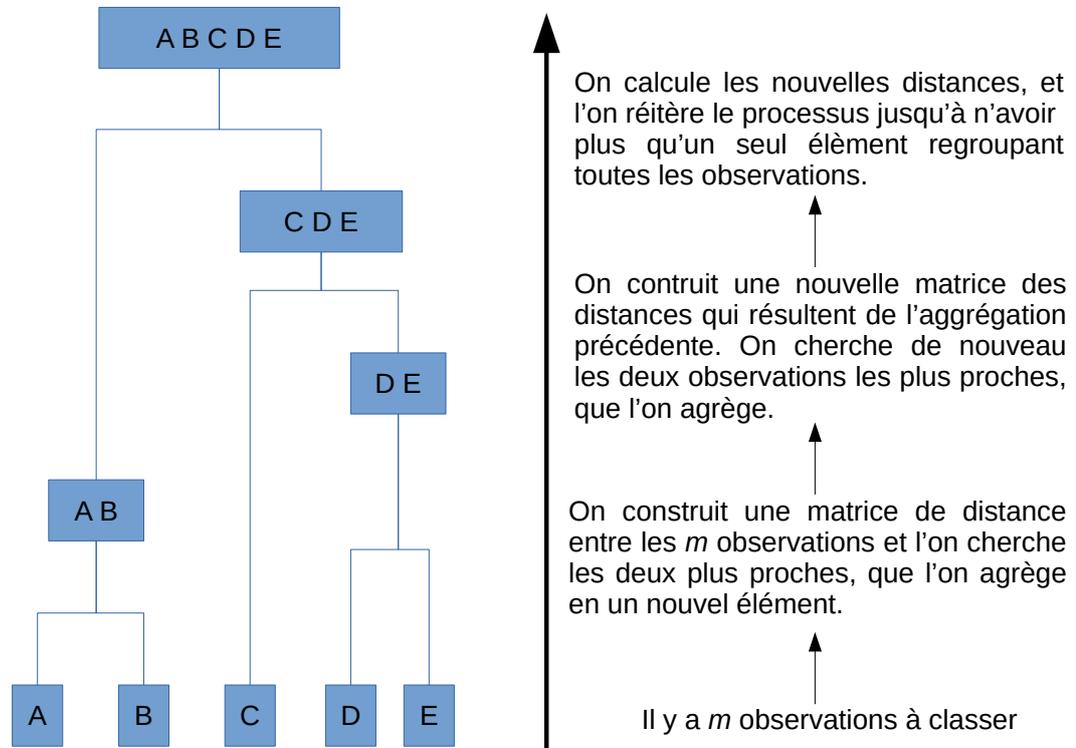


FIGURE 2.5 – Construction du dendrogramme par clustering hiérarchique ascendant.

	01-FJ048	01-PJ058	01-BD060	01-KD061	01-TC062	01-FG001	11-LA011	12-MJ012	16-BS016	19-LM019	38-SH038	39-LC039	42-CM042	47-01-MP047	57-01-CJ057	63-01-LC063
01-FJ048	0	7,76	8,14	8,38	7,07	5,67	7,25	5,07	6,51	5,43	6,41	4,39	6,06	4,18	5,21	4,63
01-PJ058	7,76	0	7,14	7,64	6,47	8,49	7,37	7,70	7,84	6,84	8,24	8,13	8,30	7,90	6,87	9,06
01-BD060	8,14	7,14	0	6,25	6,74	9,31	7,31	7,70	6,51	6,86	8,32	7,84	7,39	8,20	6,91	8,62
01-KD061	8,38	7,64	6,25	0	6,83	10,75	9,00	8,07	4,83	6,56	6,99	8,19	6,19	5,40	5,57	7,56
01-TC062	7,07	6,47	6,74	6,83	0	8,85	7,42	6,78	6,86	6,12	7,16	6,61	6,74	6,28	6,13	6,96
01-FG001	5,67	8,49	9,31	10,75	8,85	0	6,77	6,13	9,42	6,38	7,60	5,56	8,47	7,21	6,55	7,26
11-LA011	7,25	7,37	7,31	9,00	7,42	6,77	0	7,42	8,01	6,49	8,28	6,77	8,23	8,11	7,24	9,05
12-MJ012	5,07	7,70	7,70	8,07	6,78	6,13	7,42	0	6,80	5,68	6,81	5,14	6,25	5,44	6,15	5,60
16-BS016	6,51	7,84	6,51	4,83	6,86	9,42	8,01	6,80	0	5,60	7,55	7,29	6,27	6,38	6,43	7,29
19-LM019	5,43	6,84	6,86	6,56	6,12	6,38	6,49	5,68	5,60	0	5,56	5,13	5,78	5,55	4,89	5,77
38-SH038	6,41	8,24	8,32	6,99	7,16	7,60	8,28	6,81	7,55	5,56	0	5,62	7,15	5,65	5,46	6,00
39-LC039	4,39	8,13	7,84	8,19	6,61	5,56	6,77	5,14	7,29	5,13	5,62	0	6,22	4,66	5,08	4,68
42-CM042	6,06	8,30	7,39	6,19	6,74	8,47	8,23	6,25	6,27	5,78	7,15	6,22	0	5,42	6,35	5,69
47-01-MP047	4,18	7,90	8,20	5,40	6,28	7,21	8,11	5,44	6,38	5,55	5,65	4,66	5,42	0	4,74	3,40
57-01-CJ057	5,21	6,87	6,91	5,57	6,13	6,55	7,24	6,15	6,43	4,89	5,46	5,08	6,35	4,74	0	5,38
63-01-LC063	4,63	9,06	8,62	7,56	6,96	7,26	9,05	5,60	7,29	5,77	6,00	4,68	5,69	3,40	5,38	0

FIGURE 2.6 – Matrice de distance.

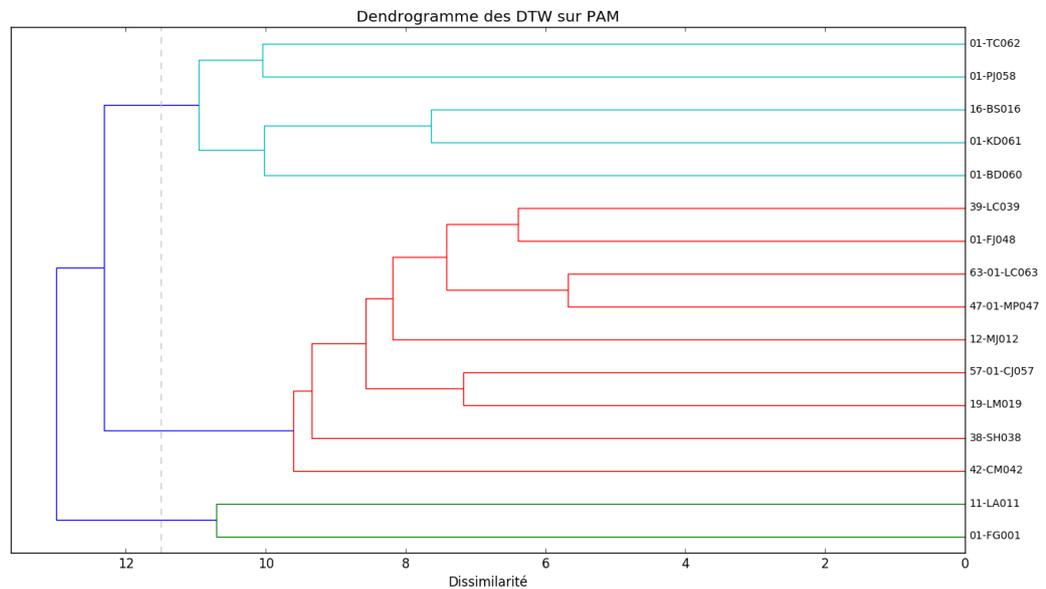


FIGURE 2.7 – Dendrogramme des signaux de la PAM, utilisant la DTW.

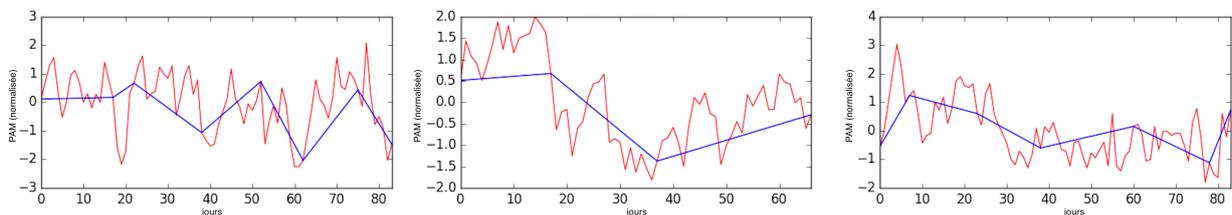


FIGURE 2.8 – PLA appliquée à nos signaux.

et non significatives) des signaux.

Il existe trois algorithmes de PLA : *sliding window*, *top-down* et *bottom-up*. Après visualisation des trois algorithmes sur nos données, nous avons choisi d'utiliser l'algorithme *sliding window*, détaillé à l'aide de l'algorithme 1, avec une marge d'erreur égale à 7.

Dans un premier temps, nous avons appliqué cette méthode sur nos signaux normalisés dont des exemples sont disponibles sur la Figure 2.8. Dans un second temps, nous avons appliqué un clustering hiérarchique à ces nouveaux signaux pour obtenir le dendrogramme de la Figure 2.9.

Algorithm 1 PLA : algorithme *sliding window*.

Avec U le signal d'origine, e la marge d'erreur fixée par l'utilisateur, et S l'ensemble de segments

```

a = 1
while a + i < taille de U do
  i = 2
  while erreur d'approximation de l'approximation linéaire de  $U[a : a + i] < e$  do
    i = i + 1
  end while
  S = concaténation de S et de l'approximation linéaire de  $U[a : a + (i - 1)]$ 
  a = a + i
end while

```

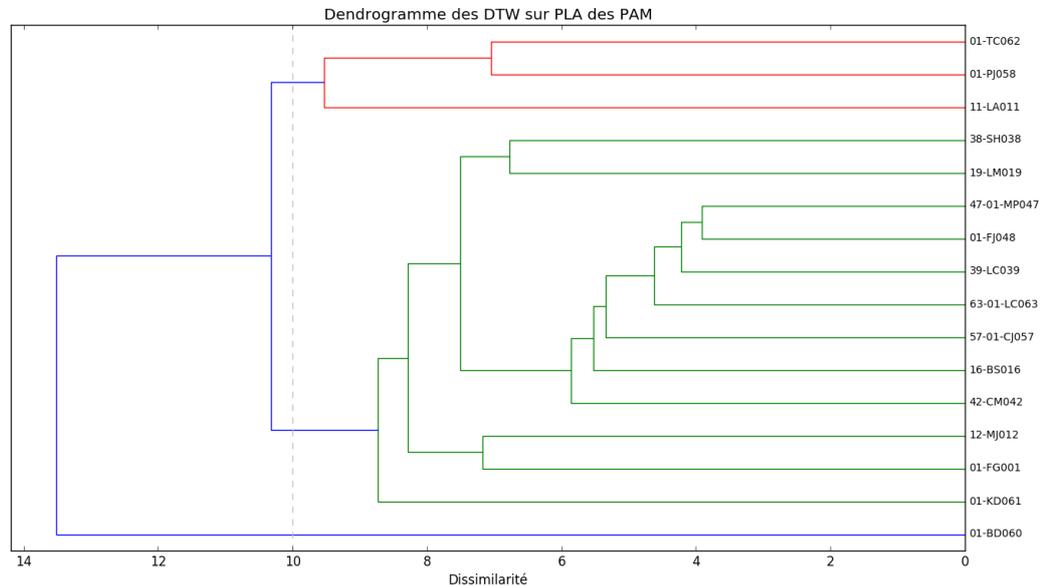


FIGURE 2.9 – Dendrogramme des signaux de la PAM, utilisant la DTW à la PLA.

2.3 Entropie

2.3.1 L'entropie de l'échantillon

Pincus propose de quantifier l'irrégularité des séries temporelles avec l'entropie approximative (*approximate entropy*) [24]. Pour palier aux limitations de l'entropie approximative, l'entropie d'échantillon (*sample entropy*) a été introduite [25]. L'entropie d'échantillon est une mesure de probabilité conditionnelle qui permet de quantifier la probabilité qu'une séquence de m points consécutifs – qui correspond à une autre séquence de même longueur – correspondra toujours à une autre séquence de longueur $m + 1$ [26]. L'entropie d'échantillon est calculée à l'aide de cette équation :

$$SampEn(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)}, \quad (2.6)$$

où r est la tolérance, m est la taille de l'échantillon, $A^m(r)$ et $B^m(r)$ sont, respectivement, la probabilité que deux séquences correspondent pour $m + 1$ et m points.

2.3.2 L'entropie multi-échelle

L'entropie mutli-échelle (*multiscale entropy* (MSE)) permet de quantifier la complexité d'une série temporelle en mesurant son irrégularité à différentes échelles temporelles [25, 26].

L'algorithme de la MSE est composé de trois étapes [27, 28] :

1. une procédure de transformation d'une série temporelle en une série temporelle en grain-grossier (*coarse-grained*) est utilisée pour dériver un ensemble de séries temporelles représentant la dynamique du système à différentes échelles temporelles. Pour un signal monovariante x de longueur N , la série temporelle en grain-grossier y^τ est obtenue à l'aide de l'équation suivante :

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad (2.7)$$

où τ est le facteur d'échelle et $1 \leq j \leq \frac{N}{\tau}$. La taille de la série temporelle en grain-grossier est égale à N/τ .

2. calcul de l'entropie de l'échantillon (Équation 2.6) pour chaque série temporelle en grain-grossier.
3. affichage de l'entropie de l'échantillon obtenue pour chaque échelle temporelle τ .

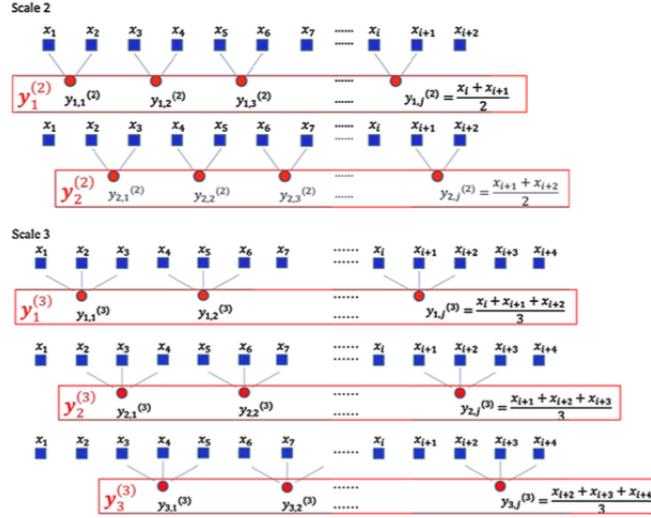


FIGURE 2.10 – RCMSE : Signal en grain grossier.
 Cette figure est extraite de la référence [29]

2.3.3 Refined composite multiscale entropy

Le calcul de la MSE génère des valeurs indéfinies pour les séries temporelles courtes [29]. Quand nous utilisons les grandes échelles temporelles τ , la série temporelle en grain-grossier admet un nombre très faible d'échantillon cela peut donc conduire à des mesures d'entropie de l'échantillon indéfinies. C'est pourquoi la *refined composite multiscale entropy* (RCMSE) a été introduite [29].

Dans la mesure de la RCMSE, la série temporelle en grain-grossier $y_k^{(\tau)}$ est obtenue pour différentes valeur de k :

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \quad (2.8)$$

où τ est le facteur d'échelle, x est le signal original, $1 \leq j \leq \frac{N}{\tau}$ et $1 \leq k \leq \tau$. Cette procédure est illustrée à l'aide de la Figure 2.10.

La mesure de la RCMSE à l'échelle τ est calculée à l'aide de l'équation suivante :

$$RCMSE(x, \tau, m, r) = -\ln \frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^m}, \quad (2.9)$$

où r est la tolérance, m la taille de l'échantillon, $n_{k,\tau}^{m+1}$ et $n_{k,\tau}^m$ sont les nombre de correspondance pour chaque τ du signal en grain-grossier pour, respectivement, $m + 1$ et m .

Pour caractériser cette mesure, nous avons choisis deux descripteurs [27, 30, 31], le premier est l'indice de complexité qui est calculé en faisant la somme des entropies des échantillons pour chaque échelle temporelle τ . Il permet de connaître le niveau de complexité de chaque signal. Le second est l'allure de chaque courbe qui permet de comparer l'évolution de l'information apportée par chaque signal à différentes échelles temporelles.

Afin de pouvoir comparer le mieux possible les données nous avons réduit notre nombre de patient à 16 en supprimant ceux qui présentaient une trop grande discontinuité dans la prise de mesure. Nous avons également réduit au même nombre de jour chaque patient afin de pouvoir faire une comparaison sur un même durée de traitement (63 jours).

Il est recommandé que la taille de l'échantillon pour calculer la RCMSE soit comprise entre 10^m et 20^m [25]. Nos données ont 63 échantillons. Nous avons donc choisi $m = 1$ et $\tau \leq 5$. De plus, nous avons fixé $r = 0,15 \times \sigma$ (où σ est l'écart-type de la série temporelle à l'échelle $\tau = 1$) [28].

Après visualisation des courbes de RCMSE de la PAM de chaque patient, nous avons observé 3 types d'allures différents : les courbes globalement décroissantes (cluster 1), les courbes globalement croissantes (cluster 2) et celles qui admettent des valeur d'entropie d'échantillon non-monotone (cluster 3). La Figure 2.11, montre la forme de la courbe de RCMSE pour chaque signal de référence de la Figure 1.2. La Figure 2.12, nous montre le résultat du clustering des courbes de RCMSE des 16 patients.

Nous avons également regroupé nos patients par le biais de l'indice de complexité, obtenus sur la RCMSE de la PAM. Le résultat de ce regroupement est illustré à l'aide de la Figure 2.13.

2.3.4 Entropie de l'échantillon croisée

L'entropie de l'échantillons croisée permet de mesurer l'asynchronisme entre deux signaux en utilisant l'entropie de l'échantillon (vue dans la partie 2.3.1)[25, 32, 33]. Pour calculer l'entropie de l'échantillon croisée il faut suivre l'algorithme 2. Nous avons dans notre cas décidé d'utiliser pour les 16 patients cette méthode sur nos deux signaux d'entrée la PAS et la PAD. Cette méthode est également adaptée aux séries temporelles courtes.

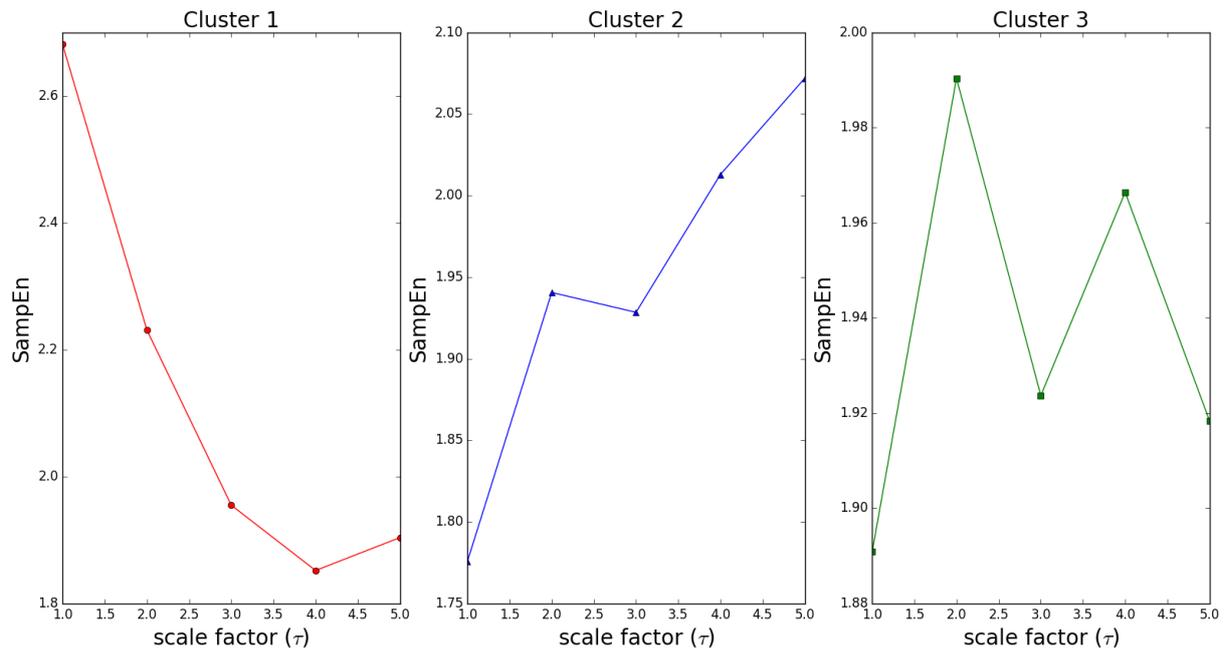


FIGURE 2.11 – RCMSE des signaux de références de la Figure 1.2.

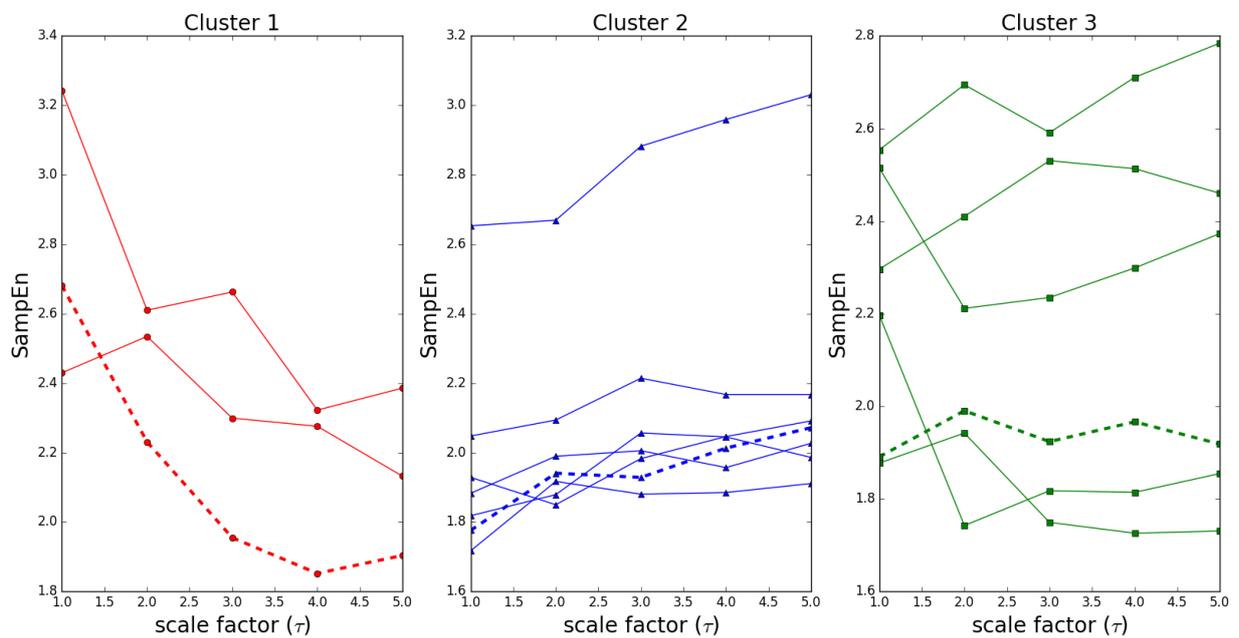


FIGURE 2.12 – RCMSE et clusters.

Le signaux originaux des courbes pointillées sont représenté à l'aide de la Figure 1.2

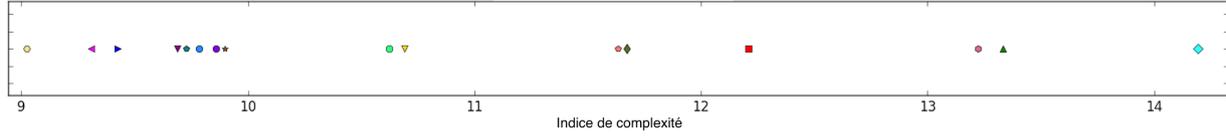


FIGURE 2.13 – Indice de complexité de tous les patients.

Algorithm 2 Calcul de l'entropie de l'échantillon croisé.

$u = (u(1), u(2), \dots, u(N))$ and $v = (v(1), v(2), \dots, v(N))$

Fixer les paramètres m et r

for all $1 \leq i \leq N - m$ **do**

$x(i) = (u(i), u(i + 1), \dots, u(i + m - 1))$

$y(i) = (v(i), v(i + 1), \dots, v(i + m - 1))$

end for

for all $i \leq N - m$ **do**

$B_i^m(r)(v||u) = \frac{\text{nombre pour } 1 \leq j \leq N - m \text{ tel que } d[x_m(i), y_m(j)] \leq r}{N - m}$

$A_i^m(r)(v||u) = \frac{\text{nombre pour } 1 \leq j \leq N - m \text{ tel que } d[x_{m+1}(i), y_{m+1}(j)] \leq r}{N - m}$

where $d[x_m(i), y_m(j)] = \max\{|u(i + k) - v(j + k)| : 0 \leq k \leq m - 1\}$

end for

$B^m(r)(v||u) = \frac{\sum_{i=1}^{N-m} B_i^m(r)(v||u)}{N - m}$

$A^m(r)(v||u) = \frac{\sum_{i=1}^{N-m} A_i^m(r)(v||u)}{N - m}$

cross-SampEn = $-\ln \frac{A^m(r)(v||u)}{B^m(r)(v||u)}$

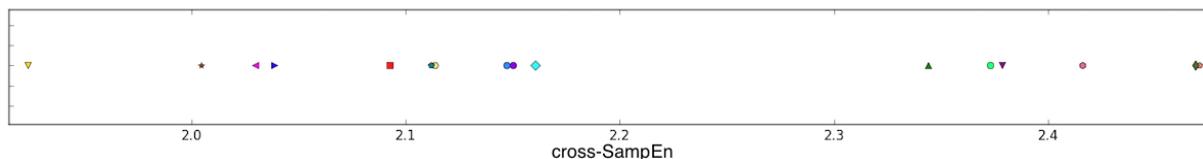


FIGURE 2.14 – Entropie de l'échantillon croisée de nos patients.

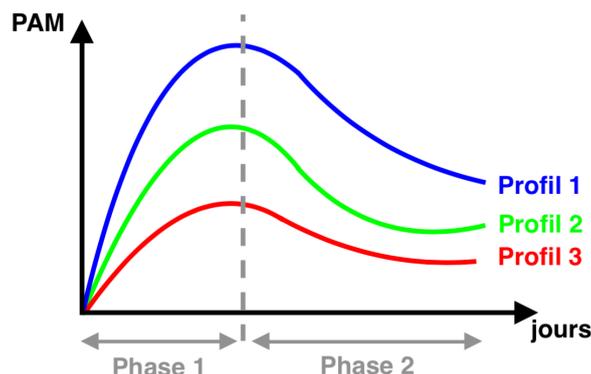


FIGURE 2.15 – Profils théoriques de l'évolution de la PAM au cours du traitement.

Dans notre cas, la PAS et la PAD ne sont pas des signaux de même amplitude et de même ordre de grandeur c'est pourquoi nous avons dû normaliser ces signaux en utilisant la formule de l'équation 2.5 de la partie 2.2.1 [34].

Le résultat de cette mesure sur nos patients est illustrée par la Figure 2.14.

2.4 Analyse de la première phase du traitement

Comme décrit dans la Figure 2.15, l'évolution de la tension pendant le traitement admet un profil qui a une première phase ascendante et une seconde décroissante. La première phase dure entre 14 et 21 jours et la seconde jusqu'à environ 30 jours. Par cette méthode, nous allons essayer de mettre en lien le nombre de dépassement des seuils de tension (vus dans la partie 1.2.1) avec l'allure des courbes de la PAM.

Pour détecter le profil des patients nous avons approximé par une droite ($y = ax + b$) la courbe de l'évolution de la PAM sur chacune des périodes ($P1a$ [0 à 13 jours], $P1b$ [14 à 20 jours], $P1$ [0 à 20 jours], $P2$ [21 à 32 jours]).

Le premier objectif a été de savoir quelle période ($P1a$ ou $P1$) nous devons prendre pour la première phase (ascendante). Nous avons mis au point l'algorithme 3.

Nous rappelons tout d'abord que l'approximation se fait par l'équation d'une droite $y = ax + b$ donc a nous donne le coefficient directeur de la droite et donc son allure (croissante ou décroissante).

Algorithm 3 Profilage des courbes de la PAM.

```

if  $a_{P1a} > 0$  and  $a_{P1b} < 0$  then
  La première phase sera égale à  $P1a$  [0 à 13 jours]
else
  if  $a_{P1a} > 0$  and  $a_{P1b} > 0$  then
    La première phase sera égale à  $P1$  [0 à 20 jours]
  end if
else
  if  $a_{P1a} < 0$  and  $a_{P1} > 0$  then
    La première phase sera égale à  $P1$  [0 à 20 jours]
  end if
else
  Le patient ne respecte pas le profil
end if

```

Ensuite nous avons calculé le taux de dépassement journalier qui revient à savoir sur combien de jours le patient a eu une PAS ou une PAD au dessus du seuil sur la première phase du traitement. Puis nous devons tracer le graphe représentant ce taux de dépassement en fonction de a .

Nous nous sommes arrêtés à 32 jours car c'est le nombre minimal d'échantillon des patients ayant un nombre d'échantillon supérieur à 30. Ce choix nous a permis d'inclure 7 nouveaux patients par rapport aux études précédentes.

Ce premier algorithme nous a fait ressortir 6 patients qui ne respectent pas le profil théorique. Ces patients sont donc situés à gauche de la ligne grise en pointillés sur la Figure 2.16.

La figure 2.16 représente le taux de dépassement de chaque patient durant la première phase en fonction du coefficient directeur de cette première phase. Nous avons ajouté en annotation bleue le taux de dépassement de chaque patient pour la seconde phase.

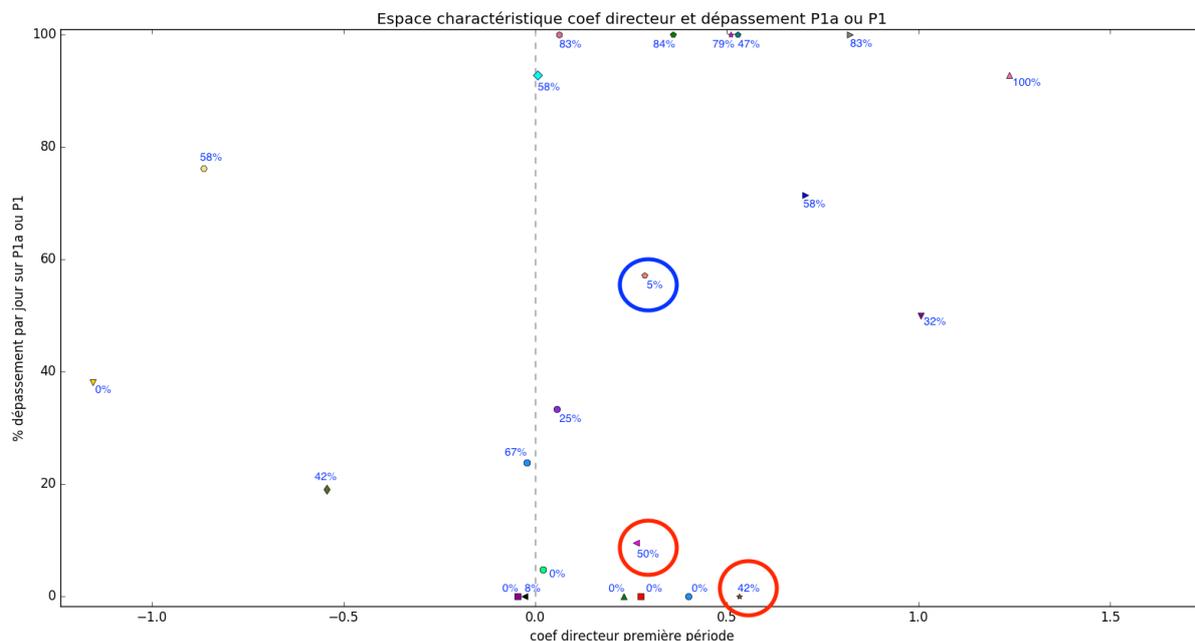


FIGURE 2.16 – Taux de dépassement de la phase 1 en fonction du coefficient directeur a .
Les % associés aux points sont les taux de dépassement en phase 2.

Nous pouvons constater que les patients qui se situent en bas du graphe ont majoritairement un taux de dépassement faible avant et après (groupe 1), et inversement pour ceux qui se situent en haut du graphe (taux de dépassement important avant et après (groupe 2)). La Figure 2.17 nous montre trois signaux du groupe 1 et la Figure 2.18 nous montre trois signaux du groupe 2.

Nous avons tout de même des problèmes avec 2 patients (entourés en rouge) qui se trouvent en bas du graphe alors qu'ils ont un taux de dépassement qui augmente significativement entre la phase 1 et la phase 2. Nous avons également un autre patient qui pose problème (entouré en bleu) qui se trouve au milieu, voir plutôt haut, alors que son taux de dépassement diminue fortement. Pour les patients entourés en rouge, représentés en bleu et en rouge sur la Figure 2.19, les profils sont vraiment différents car pour le bleu le taux de dépassement sur le long terme tend vers 0 alors que pour le rouge il est élevé. Il existe une différence dans le traitement de ces deux patients (prise ou non d'un anticoagulant), elle pourrait être une réponse à ce problème de classification. Pour le patient entouré en bleu, représenté en vert sur la Figure 2.19, le profil montre que le taux de dépassement est élevé sur le long terme et donc rend cohérent son placement dans l'espace caractéristique.

Suite à cette analyse de la première phase de traitement nous pourrions définir la

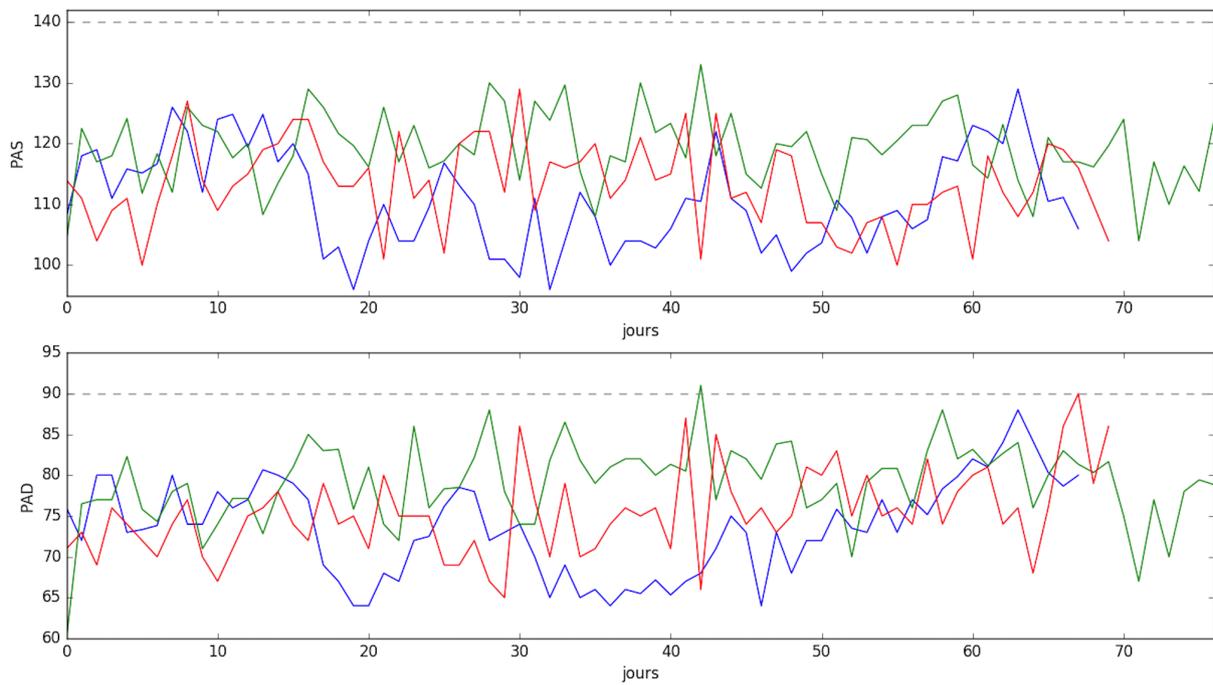


FIGURE 2.17 – Groupe 1 de l'analyse de la première phase du traitement.

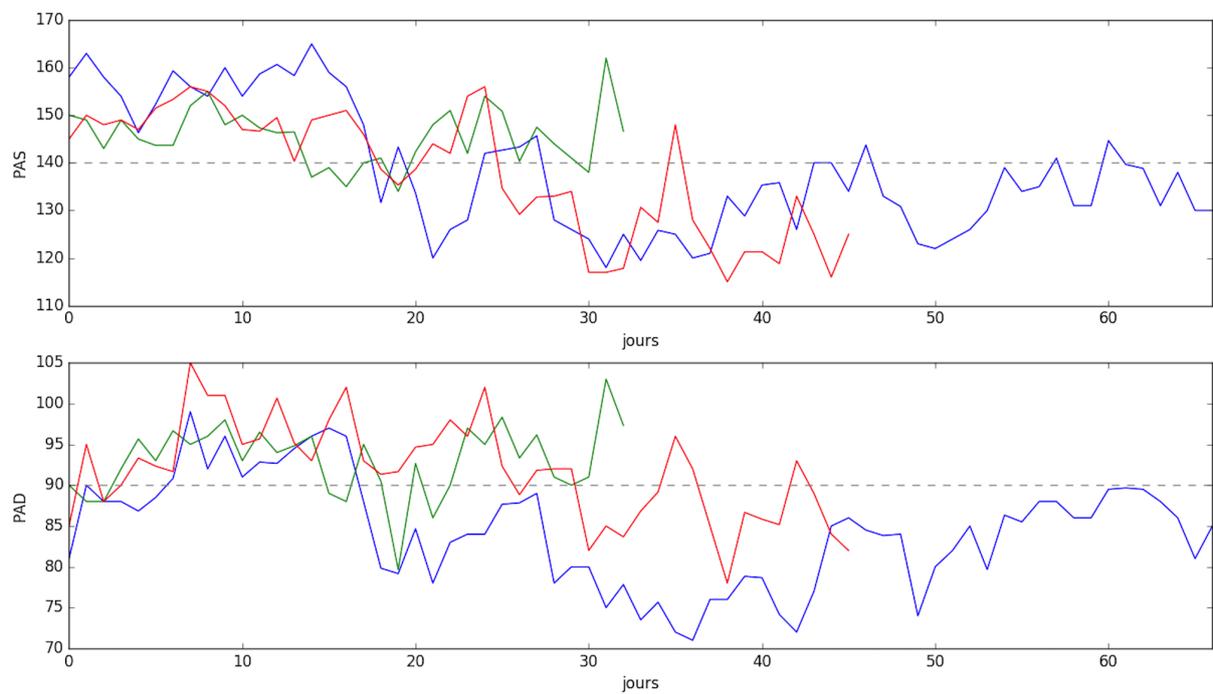


FIGURE 2.18 – Groupe 2 de l'analyse de la première phase du traitement.

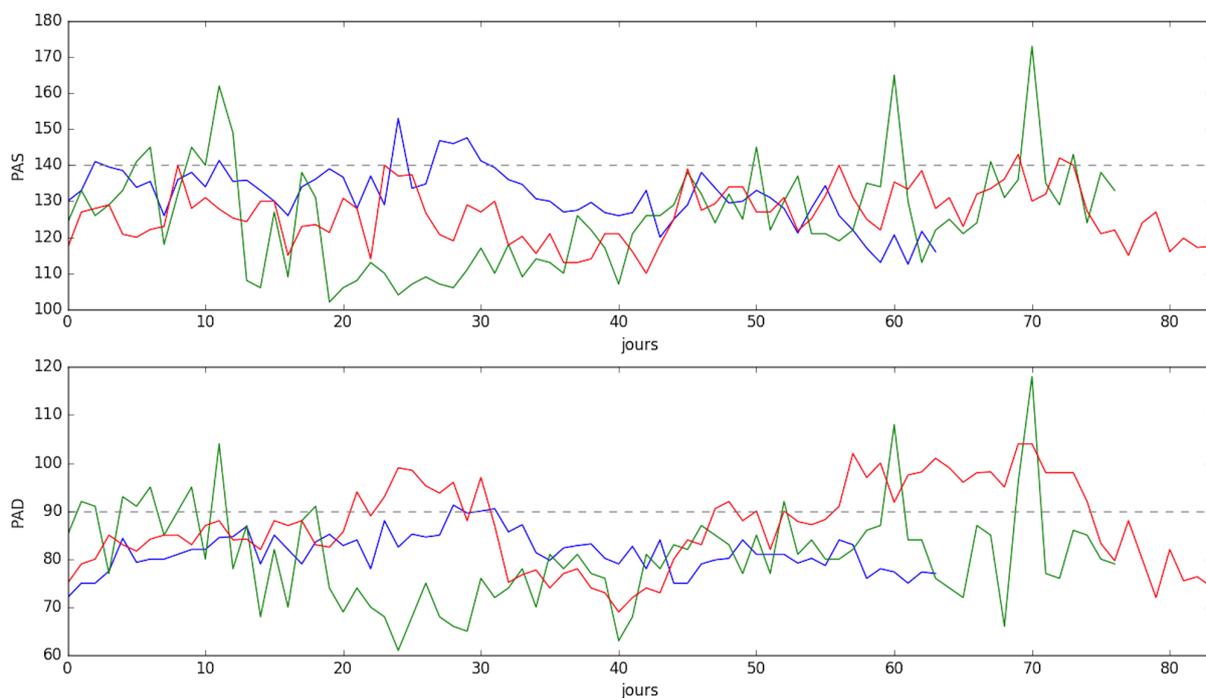


FIGURE 2.19 – Profils des patient qui nous posent problème.

stratégie de télé-monitoring suivante : tous les patients dont le taux de dépassement en phase 1 est supérieur à 15% doivent continuer à être suivis par tensiomètre connecté, il est de même pour les patients qui ne respectent pas le profil de la première phase du traitement. Cet algorithme d'aide à la décision nous a permis donc d'éliminer 6 patients du processus de télé-monitoring.

2.5 Résultats

Chaque méthode, présentée dans les sections précédentes, a donné un clustering différent. La méthode par RCMSE a été confrontée à un diagnostic médical qui nous a montré qu'il existait un lien entre le traitement antihypertenseur et les clusters obtenus. Il sera donc intéressant par la suite de confronter les autres méthodes à une analyse médicale des patients pour juger la pertinence ou non de chaque méthode. Pour chaque autre méthode, le résultat a été comparé à celui obtenu grâce à la RCMSE, en considérant 3 clusters. Le Tableau 2.1 est un récapitulatif du clustering obtenu par méthode.

Dossier	RCMSE – pente	RCMSE – IC	CSE	DTW	PLA-DTW	DS 3D	DS 4D
01-FJ048	B	C	B	A	A	C	B
01-PJ058	C	C	A	C	B	A	C
01-BD060	C	B	A	C	C	A	A
01-KD061	B	C	C	C	A	A	A
01-TC062	C	C	C	C	B	B	C
01-FG001	A	A	A	B	A	A	A
11-LA011	C	A	B	B	B	A	B
12-MJ012	C	A	A	A	A	B	A
16-BS016	B	C	C	C	A	C	A
19-LM019	A	B	A	A	A	B	C
38-SH038	C	C	B	A	A	C	C
39-LC039	B	B	B	A	A	C	C
42-CM042	A	A	A	A	A	C	A
47-01-MP047	B	C	B	A	A	B	C
57-01-CJ057	B	C	B	A	A	C	A
63-01-LC063	B	A	C	A	A	B	A

TABLE 2.1 – Récapitulatif des clustering par méthode.

CSE : cross-SampEn | IC : Indice de Complexité | DS : Descripteurs Statistiques

Conclusion et perspectives

Cette étude nous a permis de passer en revue plusieurs méthodes de traitement de série temporelle. La RCMSE a pour sa part été confrontée à un diagnostic médical qui a révélé qu'il existe un lien entre l'allure de la RCMSE des signaux et la prise d'antihypertenseur. Cette prescription médicale semble donc influencer la complexité de la pression artérielle. Ce résultat nous a permis d'écrire un article (disponible en annexe) que nous avons soumis pour participer à une conférence internationale sur les objets connectés dans le domaine de la santé.

Pour la suite de notre étude, nous pourrions confronter les résultats des autres méthodes à un diagnostic médical. Ceci nous permettrait de voir si il existe d'autres liens entre les caractéristiques du signal de la pression artérielle et les caractéristiques médicales de chaque patient. L'objectif futur est que cette analyse permette de mettre en place un algorithme de reconnaissance de profil assurant une aide à la prise de décision pour les médecins.

D'un point de vue personnel, ce stage m'a permis de découvrir le monde de la recherche ainsi que l'analyse de séries temporelles. Cette expérience a conforté mon choix de continuer en doctorat, dans le domaine du traitement de signaux médicaux, dès l'année prochaine. Ce projet m'a permis d'acquérir les bases en traitement du signal utiles pour mener à bien le sujet qui me sera confié pendant mes trois années de thèse.

Bibliographie

- [1] “World cancer research fund international.” <http://www.wcrf.org>, consulté en mai 2017.
- [2] “Le magazine des objets connectés et innovants.” <http://www.objetconnecte.net>, consulté en juin 2017.
- [3] A. Whitmore, A. Agarwal, and L. Da Xu, “The Internet of Things—A survey of topics and trends,” *Information Systems Frontiers*, vol. 17, no. 2, pp. 261–274, 2015.
- [4] N. Seydoux, M. Ben Alaya, N. Hernandez, T. Monteil, and O. Haemmerlé, “Sémantique et Internet des objets : d’un état de l’art à une ontologie modulaire,” in *26es Journées francophones d’Ingénierie des Connaissances*, (Rennes, France), 2015.
- [5] Y. Yin, Y. Zeng, X. Chen, and Y. Fan, “The internet of things in healthcare : An overview,” *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, 2016.
- [6] F.-A. Allaert and N.-J. Mazen, “Évaluation des objets connectés et des applications de santé,” *Actualités Pharmaceutiques*, vol. 55, no. 556, pp. 29–32, 2016.
- [7] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. S. Kwak, “The Internet of Things for Health Care : A Comprehensive Survey,” *IEEE Access*, vol. 3, pp. 678–708, 2015.
- [8] E. Borgia, “The Internet of Things vision : Key features, applications and open issues,” *Computer Communications*, vol. 54, pp. 1–31, 2014.
- [9] M. Pavel, H. B. Jimison, I. Korhonen, C. M. Gordon, and N. Saranummi, “Behavioral Informatics and Computational Modeling in Support of Proactive Health Management and Care,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 12, pp. 2763–2775, 2015.
- [10] B. Brouard, P. Bardo, M. Vignot, C. Bonnet, and S. Vignot, “E-santé et m-santé : état des lieux en 2014 et apports potentiels en oncologie,” *Bulletin du Cancer*, vol. 101, no. 10, pp. 940–950, 2014.
- [11] F. Denis, S. Yossi, A.-L. Septans, A. Charron, E. Voog, O. Dupuis, G. Ganem, Y. Pointreau, and C. Letellier, “Improving Survival in Patients Treated for a Lung Cancer Using Self-Evaluated Symptoms Reported Through a Web Application.,” *American Journal of Clinical Oncology*, vol. Publish Ahead of Print, 2015.

- [12] X. Rafael-Palou, A. Steblin, and E. Vargiu, “Remotely Supporting Patients with Obstructive Sleep Apnea at Home,” in *Internet of Things Technologies for HealthCare : Third International Conference, HealthyIoT 2016, Västerås, Sweden, October 18-19, 2016, Revised Selected Papers*, pp. 119–124, Cham : Springer International Publishing, 2016.
- [13] A. Jamin, J.-B. Fasquel, M. Lhommeau, E. Cornet, S. Abadie-Lacourtoisie, S. Henni, and G. Leftheriotis, “An Aggregation Platform for IoT-Based Healthcare : Illustration for Bioimpedancemetry, Temperature and Fatigue Level Monitoring,” in *Internet of Things Technologies for HealthCare : Third International Conference, HealthyIoT 2016, Västerås, Sweden, October 18-19, 2016, Revised Selected Papers*, pp. 125–130, Cham : Springer International Publishing, 2016.
- [14] E. Basch, A. M. Deal, M. G. Kris, H. I. Scher, C. A. Hudis, P. Sabbatini, L. Rongak, A. V. Bennett, A. C. Dueck, T. M. Atkinson, J. F. Chou, D. Dulko, L. Sit, A. Barz, P. Novotny, M. Fruscione, J. A. Sloan, and D. Schrag, “Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment : A Randomized Controlled Trial,” *Journal of Clinical Oncology*, vol. 34, no. 6, pp. 557–565, 2016.
- [15] S. Rani and G. Sikka, “Recent Techniques of Clustering of Time Series Data : A Survey,” *International Journal of Computer Applications*, vol. 52, no. 15, pp. 1–9, 2012.
- [16] S. Aghabozorgi, A. S. Shirkorshidi, and T. Y. Wah, “Time-series clustering – A decade review,” *Information Systems*, vol. 53, pp. 16–38, 2015.
- [17] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, “Feature-based classification of time-series data,” in *Information processing and technology*, pp. 49–61, Nova Science Publishers, Inc., 2001.
- [18] U. Mori, A. Mendiburu, and J. A. Lozano, “Similarity Measure Selection for Clustering Time Series Databases,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 1, pp. 181–195, 2016.
- [19] P. Senin, “Dynamic time warping algorithm review,” *CSDL Technical report*, 2008.
- [20] E. Biernat and M. Lutz, *Data science : fondamentaux et études de cas – Machine learning avec Python et R*. Blanche, Eyrolles, 2015.
- [21] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.
- [22] H. Li, C. Guo, and W. Qiu, “Similarity measure based on piecewise linear approximation and derivative dynamic time warping for time series mining,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 14732–14743, 2011.
- [23] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in *Proceedings 2001 IEEE International Conference on Data Mining*, pp. 289–296, 2001.

- [24] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [25] J. S. Richman and J. Randall Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, p. H2039, 2000.
- [26] A. Humeau-Heurtier, "The Multiscale Entropy Algorithm and Its Variants : A Review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, 2015.
- [27] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Physical Review Letters*, vol. 89, no. 6, p. 068102, 2002.
- [28] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, p. 021906, 2005.
- [29] S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, and C.-K. Peng, "Analysis of complex time series using refined composite multiscale entropy," *Physics Letters A*, vol. 378, no. 20, pp. 1369–1374, 2014.
- [30] Y.-C. Chang, H.-T. Wu, H.-R. Chen, A.-B. Liu, J.-J. Yeh, M.-T. Lo, J.-H. Tsao, C.-J. Tang, I.-T. Tsai, and C.-K. Sun, "Application of a Modified Entropy Computational Method in Assessing the Complexity of Pulse Wave Velocity Signals in Healthy and Diabetic Subjects," *Entropy*, vol. 16, no. 7, pp. 4032–4043, 2014.
- [31] D.-R. Kong and H.-B. Xie, "Use of modified sample entropy measurement to classify ventricular tachycardia and fibrillation," *Measurement*, vol. 44, no. 4, pp. 653–662, 2011.
- [32] T. Zhang, Z. Yang, and J. H. Coote, "Cross-sample entropy statistic as a measure of complexity and regularity of renal sympathetic nerve activity in the rat.," *Experimental physiology*, vol. 92, no. 4, pp. 659–669, 2007.
- [33] W. Shi and P. Shang, "Cross-sample entropy statistic as a measure of synchronism and cross-correlation of stock markets," *Nonlinear Dynamics*, vol. 71, no. 3, pp. 539–554, 2013.
- [34] C. C. Chiu, S. J. Yeh, and T. Y. Li, "Assessment of Diabetics with Various Degrees of Autonomic Neuropathy Based on Cross-Approximate Entropy," in *5th Kuala Lumpur International Conference on Biomedical Engineering 2011 : (BIOMED 2011) 20-23 June 2011, Kuala Lumpur, Malaysia*, pp. 266–269, Berlin, Heidelberg : Springer Berlin Heidelberg, 2011. DOI : 10.1007/978-3-642-21729-6_70.

Annexe

Article Conférence Healthy-IoT 2017

Cet article a été soumis le 23/05/2017 pour la 4ème Conférence Internationale de l'EAI¹ sur les technologies IoT dans le domaine de la santé accueillie par le LARIS le 24 et 25 octobre 2017². La réponse pour savoir si cet article est retenu nous sera communiquée en septembre 2017.

1. European Alliance for Innovation
2. <http://healthyiot.org/2017/show/home>

Characterization of home-acquired blood pressure time series using multiscale entropy for patients treated against kidney cancer

Antoine JAMIN¹, Jean-Baptiste FASQUEL¹, Anne HUMEAU-HEURTIER¹,
Pierre ABRAHAM², Georges LEFTHERIOTIS³, and Samir HENNI^{1,2}

¹ LARIS, Université d'Angers, 62 Avenue Notre-Dame du Lac, 49000 Angers, France

² University Hospital Center of Angers, Vascular Departement and MITOVASC,
Angers, France

³ University Hospital Center of Nice, Vascular Departement, Nice, France

Abstract. This study deals with the telemonitoring, with a connected tensiometer, of 16 patients treated for a kidney cancer. Each one of these patients recorded his/her blood pressure at home during 63 days and the data was sent to his/her medical doctor. At the same time they were treated with antihypertensive medication when necessary. In this work, our goal was to analyze the complexity of the blood pressure time series. For this purpose, we proposed to use the refined composite multiscale entropy (RCMSE) measures. Our results show that the patterns of RCMSE through temporal scales evolve with the antihypertensive medication. The later might therefore have an impact on home-acquired blood pressure complexity. RCMSE could therefore be an interesting information theory-based tool to study home-acquired physiological data.

Key words: telemonitoring, connected tensiometer, blood pressure, time series, multiscale entropy, clustering, irregularity, complexity

1 Introduction

Kidney cancer is the 12th kind of cancer in terms of frequency in the world. It represents 338,000 new cases diagnosed in 2012 [1]. It can be treated by a VEGF¹ chemotherapy that consists in eliminating the capillaries of the tumor. However such a treatment can lead to blood pressure increases. This is why antihypertensive medication are often given to the patients. Our work deals with patients treated against a kidney cancer using a VEGF chemotherapy. During the treatment, patients were daily monitored at home using a connected tensiometer (see Figure 1). Such an IoT-based sensor [2, 3] facilitates the following of blood pressure increases. Our goal herein is to study, with the refined composite multiscale entropy (RCMSE), the complexity of the blood pressure time series.

¹ Vascular Endothelial Growth Factor

RCMSE is an improved version of the multiscale entropy (MSE). MSE is based on the use of sample entropy, a measure of time series irregularity, at different time scales [4].

The paper is organized as follows: the MSE and RCMSE algorithms and the measurement procedure are introduced in Section 2. Our results are then detailed and discussed in Section 3. The paper ends with a Conclusion.

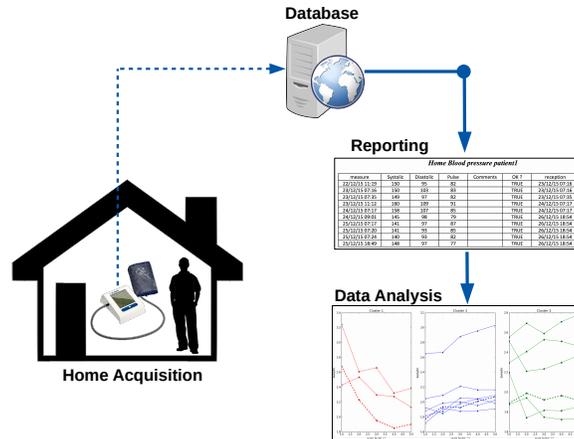


Fig. 1: Project overview

2 Materials and Methods

2.1 Sample Entropy

Pincus proposed to quantify the irregularity of time series with approximate entropy (ApEn) [5]. To overcome the limitation of ApEn, sample entropy has later been introduced [6]. Sample entropy is a conditional probability measure that quantifies the likelihood that a sequence of m consecutive data points – that matches another sequence of the same length – will still match the other sequence when their length is increased by one sample ($m+1$) [4]. Sample entropy is computed as:

$$SampEn(m, r, N) = -\ln \frac{A^m(r)}{B^m(r)}, \quad (1)$$

where r is the tolerance, m is the sample length, $A^m(r)$ and $B^m(r)$ are, respectively, the probability that two sequences will match for $m+1$ and m points.

2.2 Multiscale entropy

MSE allows to quantify the complexity of time series by measuring its irregularity at different time scales [4, 6].

The MSE algorithm is composed of three steps [7, 8]

1. a coarse-graining procedure is used to derive a set of time series representing the system dynamics on different time scales. For a monovariate discrete signal x of length N , the coarse-grained time series y^τ is computed as:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad (2)$$

where τ is the scale factor and $1 \leq j \leq \frac{N}{\tau}$. The length of the coarse-grained time series is N/τ .

2. computation of the sample entropy for each coarse-grained times series.
3. plot of the sample entropy for each time scale τ .

2.3 Refined composite multiscale entropy

MSE generates some undefined values for short time series [9]. When large scale factors τ are used, the coarse-grained time series may have a small number of samples. This may lead to undefined sample entropy values. This is why RCMSE has been introduced [9].

In RCMSE, the coarse-grained signal $y_k^{(\tau)}$ is computed for different values of a parameter k :

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \quad (3)$$

where τ is the scale factor, x is the original signal, $1 \leq j \leq \frac{N}{\tau}$ and $1 \leq k \leq \tau$.

The RCMSE at scale τ is calculated using the following formulation:

$$RCMSE(x, \tau, m, r) = -\ln \frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^m}, \quad (4)$$

where r is the tolerance, m is the sample length, $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^m$ are the number of matched vector pairs (computed on $y_k^{(\tau)}$) for $m+1$ and m , respectively.

2.4 Measurement procedure

The study was conducted on 16 patients (5 women, 11 men; $171 \text{ cm} \pm 9 \text{ cm}$; $87 \text{ kg} \pm 29 \text{ kg}$; $62 \text{ years} \pm 8 \text{ years}$). The 16 patients were daily monitored, during 63 days, using a connected tensiometer [10]. Each recorded time series (Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) time series) had therefore 63 samples. In our work, the mean blood pressure (MBP), defined as:

$$MBP = \frac{SBP + 2 \times DBP}{3}, \quad (5)$$

was processed by RCMSE.

2.5 Parameters used

To compute RCMSE, 3 parameters have to be set : m , r , and τ . It has been recommended that, to compute the sample entropy, the time series length has to be between 10^m and 20^m [6]. Our data have 63 samples. We therefore have chosen $m = 1$ and time scale $\tau \leq 5$. Moreover, we have chosen $r = 0.15 \times \sigma$ (where σ is the standard deviation of time series at scale factor $\tau = 1$) [8].

3 Results and Discussion

Figure 2 presents three examples of MBP data. Figure 3 shows the corresponding RCMSE curves. By analyzing all the time series we observe 3 kinds of RCMSE patterns : global decreasing sample entropy with scales (cluster 1), global increasing sample entropy with scales (cluster 2), and non-monotonic sample entropy values with scales (cluster 3); see Figure 4.

If sample entropy increases with scales, this means that the signal contains complex structures across multiple scales. If the sample entropy decreases with scales, this means that the signal has information only on the shortest scale (similarly to white noise [8]). In the undetermined case, no definite conclusion can be drawn.

We can observe that the signals of cluster 1 are similar to RCMSE of a white noise: the signal irregularity decreases with scale. Cluster 2 corresponds to an increase of the signal irregularity with scales. Cluster 3 corresponds to an intermediate situation.

These clusters have been compared with the number of hypertensive medication taken by the patients. As reported in Table 1, for cluster 1, patients received only 1 antihypertensive treatment. For cluster 2, patients received 2 antihypertensive treatments. For cluster 3, patients received 3 antihypertensive treatments.

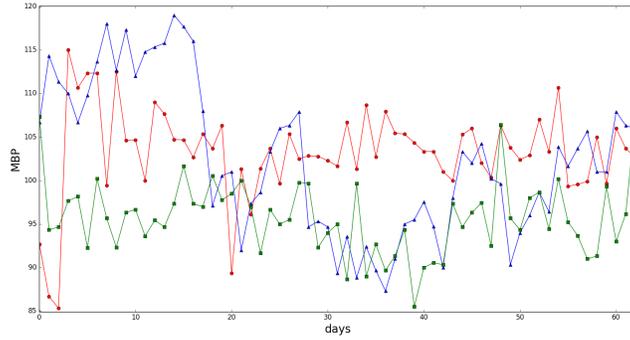


Fig. 2: Representative signals for each cluster (cluster 1 : red — cluster 2 : blue — cluster 3 : green)

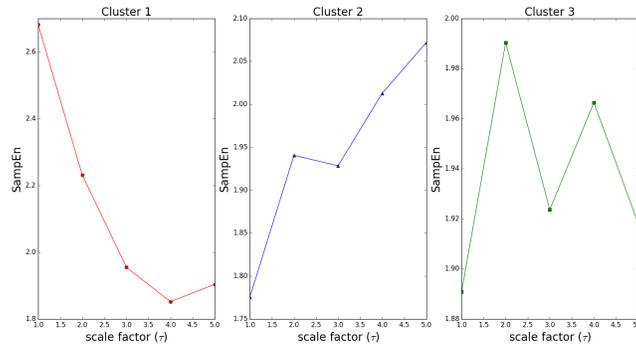


Fig. 3: RCMSE and clusters of representative signals shown in Figure 2.

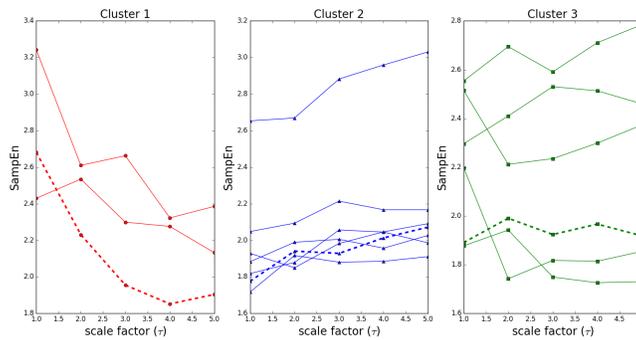


Fig. 4: RCMSE and clusters. Original signals leading to dotted RCMSE are represented in Figure 2

Cluster	Cluster 1	Cluster 2	Cluster 3
RCMSE shape	decrease	increase	undetermined
#patients	3	7	6
#medication	1	2	3

Table 1: Clusters, RCMSE and #medication

Anti-hypertensive medication may therefore play a role in the complexity of mean blood pressure time series.

4 Conclusion

In this work we studied the complexity of blood pressure time series with RCMSE. Our work shows that antihypertensive medication might have an influence on mean blood pressure complexity. Our study shows that the complexity of data extracted by connected devices might have an interest for physiological purposes. RCMSE could now be applied to other IoT-based physiological data. Next step of this work will consist in identifying patient profiles in order to customize the telemonitoring.

References

1. World Cancer Research Fund International, <http://www.wcrf.org> [may 2017]
2. Y. Yin, Y. Zeng, X. Chen, Y. Fan: The internet of things in healthcare: An overview. *Journal of Industrial Information Integration*, Volume 1, 3–13 (2016).
3. S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, K. S. Kwak: The Internet of Things for Health Care: A Comprehensive Survey. *IEEE Access*, Volume 3, 678–708 (2015).
4. A. Humeau-Heurtier: The Multiscale Entropy Algorithm and Its Variants: A Review. *Entropy*, Volume 17, 3110–3123 (2015)
5. S. M. Pincus: Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, Volume 88, 2297–2301 (1991).
6. J. S. Richman, J. R. Moorman: Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology*, Volume 278, H2039–H2049 (2000).
7. M. Costa, A. L. Goldberger, C.-K. Peng: Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*, Volume 89, 068102 (2002).
8. M. Costa, A. L. Goldberger, C.-K. Peng: Multiscale entropy analysis of biological signals. *Physical Review E*, Volume 71, 021906 (2005).
9. S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, C.-K. Peng: Analysis of complex time series using refined composite multiscale entropy. *Physics Letters A*, Volume 378, 1369–1374 (2014).
10. Blood Pressure Monitoring - I.E.M. GmbH, <http://www.iem.de/en/health-management/blood-pressure-monitoring.html> [may 2017]

Résumé — Notre projet concerne le suivi de la pression artérielle à domicile de 28 patients, atteints de cancer du rein, par le biais d'un tensiomètre connecté. L'objectif de ce travail est d'analyser les données récupérées par le tensiomètre connecté pour pouvoir les mettre en lien avec le diagnostic médical. Pour se faire nous passerons en revues plusieurs méthodes de traitements de séries temporelles : une mesure de dissymétrie à l'aide de descripteurs statistiques, une mesure de dissimilarité par mesure de distance avec ou sans filtrage par approximation linéaire fait au préalable, une mesure de complexité par entropie multi-échelle. Nous ferons également une analyse de la première phase du traitement en étudiant le taux de dépassement de seuil de la pression artérielle. Dans ce rapport nous verrons dans un premier temps le projet et son contexte et dans un second temps nous présenterons chaque méthode de traitement de séries temporelles ainsi que les résultats obtenus.

Mots clés : cancer du rein, pression artérielle, objet connecté, analyse de séries temporelles, clustering, complexité, dissimilarité, dissymétrie.

Abstract — This study deals with the telemonitoring with a connected tensiometer, of 28 patients treated for a kidney cancer. Our goal is to link the medical diagnosis with the analysis of the data collected by the connected tensiometer. For this purpose we decide to use some different methods related to time series analysis : a dissymmetry measure with statistical descriptors, a dissimilarity measure using dynamic time warping (with and without preliminary piecewise linear approximation), an entropy-based measure with the multiscale entropy. We study also the first phase of the treatment with the blood pressure exceedance rate. In this dissertation we present the context first, and then one focuses on methods and results.

Keywords : kidney cancer, blood pressure, IoT, time series analysis, clustering, complexity, dissimilarity, dissymmetry.
