



HAL
open science

Fouille de données et segmentation de chroniques par extrema: considérations préliminaires

Michel Fliess, Cédric Join

► **To cite this version:**

Michel Fliess, Cédric Join. Fouille de données et segmentation de chroniques par extrema: considérations préliminaires. 13ème Conférence Internationale de Modélisation, Optimisation et Simulation, MOSIM 2020, Nov 2020, Agadir, Maroc. hal-02929875

HAL Id: hal-02929875

<https://polytechnique.hal.science/hal-02929875>

Submitted on 3 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from 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.

Fouille de données et segmentation de chroniques par extrema : considérations préliminaires

Data mining and time series segmentation via extrema: preliminary investigations

Michel Fliess

Cédric Join

LIX (CNRS, UMR 7161)

CRAN (CNRS, UMR 7039)

École polytechnique

Université de Lorraine, BP 239

91128 Palaiseau - France

54506 Vandœuvre-lès-Nancy - France

Michel.Fliess@polytechnique.edu

Cedric.Join@univ-lorraine.fr

AL.I.E.N.

7 rue Maurice Barrès

54330 Vézelize - France

{michel.fliess, cedric.join}@alien-sas.com

RÉSUMÉ : *La segmentation des chroniques est l'un des outils de fouilles des données. On l'aborde, ici, en prenant des extrema locaux pour points d'importance perceptuelle (PIP). Un théorème de décomposition additive des chroniques, dû à Cartier et Perrin, et des techniques algébriques d'estimation des dérivées, déjà utiles en automatique et signal, permettent d'évacuer le brouillage dû aux fluctuations rapides de toute chronique. Des illustrations numériques valident notre approche et soulignent l'importance du choix du seuil de détection.*

ABSTRACT: *Time series segmentation is one of the many data mining tools. We take here local extrema as perceptually interesting points (PIPs). The blurring of those PIPs by the quick fluctuations around any time series is treated via an additive decomposition theorem, due to Cartier and Perrin, and algebraic estimation techniques, which are already useful in automatic control and signal processing. Our approach is validated by several computer illustrations. They underline the importance of the choice of a threshold for the extrema detection.*

MOTS-CLÉS : *chroniques, segmentation, point d'importance perceptuelle (PIP), estimation algébrique.*

KEYWORDS: *time series, segmentation, perceptually important points (PIPs), algebraic estimation.*

1 Introduction

Les chroniques, ou « time series », sont omniprésentes dans la fouille de données, ou « data mining ». Parmi les multiples méthodes utilisées, on privilégie ici la *segmentation* (voir, par exemple, Abonyi et Feil (2007); Cho et Fryzlewicz (2012); Durán-Rosal et al. (2018a,b, 2019); Esling et Agon (2012); Keogh et al. (2004); Kim et al. (2019); Lee et al. (2018); Liu et al. (2008); Lu et Huang (2020); Martí et al. (2014)), c'est-à-dire une approximation par un nombre fini de segments de droite. De nombreuses études concrètes l'ont déjà utilisée dans des domaines variés : hydrométéorologie (Hubert et al. (1989)), économie (Cheong et al. (2012)), finance (Wan et al. (2016); Wan et Si (2017)), mouvement animal (Edelhoff et al. (2016)), vision par ordinateur (Liu et al. (2019)), efficacité énergétique des machines outils (Seevers et al. (2019)), collecte de données par drones aquatiques (Castellini et al. (2020)), ... Prendre pour bouts des segments des *points d'importance perceptuelle (PIP)*, ou « *perceptually important points* », consolide ce point de vue (voir, par exemple, Fu (2011); Fu et al. (2017)), qui

l'est encore davantage si ces PIP sont des extrema locaux (Pratt et Fink (2002); Fink et Pratt (2004); Fink et Gandhi (2011); Yin et al. (2011)). La détection des PIP est brouillée par les fluctuations rapides de toute chronique. Comment isoler des extrema locaux sur la ligne bleue de la figure 1 ? Les publications existantes ignorent, semble-t-il, cette difficulté majeure. Cet article présente un cadre mathématique et algorithmique afin de détecter ces extrema locaux en dépit des oscillations. Quelques exemples académiques sur la réduction de dimension, préalable à toute fouille de données, illustrent notre propos.

Notre approche des chroniques, qui s'appuie sur (Fliess et al. (2018)), consacré à la prévision et à la gestion du risque pour l'énergie photovoltaïque, est née en ingénierie financière (Fliess et Join (2009); Fliess et al. (2011)). Elle a été utilisée pour la prévision du trafic autoroutier (Abouaïssa et al. (2016)) et des ressources nécessaires en informatique nuagique, ou « cloud computing » (Fliess et al. (2019)). Le pivot en est un théorème dû à Cartier et Perrin (1995) : toute chronique X satisfaisant une hypothèse très

lâche, s'écrit $X = E(x) + X_{\text{fluctuat}}$, où $E(x)$ est la *moyenne*, ou *tendance* (« trend »), et X_{fluctuat} les fluctuations rapides autour de 0. La ligne rouge de la figure 1 correspond à la moyenne. Le théorème de Cartier et Perrin s'exprime dans le langage de l'*analyse non standard*, introduit par Robinson (1996) et outil merveilleux pour l'intuition (voir, par exemple, Lobry et Sari (2008)). On ne recherche pas les extrema sur X , mais sur $E(X)$, plus lisse. La détection d'un extremum est liée à l'annulation de la dérivée d'ordre 1. On y parvient grâce aux techniques d'*estimation algébrique* (Fliess et al. (2008); Mboup et al. (2009)) pour traiter du cas bruité, connu comme important et difficile (Lanczos (1956)). Si la dérivée est proche de 0 durant un certain temps, la chronique est remplacée par un plateau, c'est-à-dire un segment horizontal. Des techniques algébriques analogues ont déjà été employées pour détecter des ruptures, ou « change-points » (Fliess et al. (2010)).

Cette communication est organisée comme suit. Les rappels du § 2 sont consacrés aux chroniques et à l'estimation algébrique. On présente au § 3 des illustrations numériques, qui soulignent l'importance du seuil de détection des extrema. Quelques prolongements sont proposés au § 4.

2 Rappels sur les chroniques et l'estimation algébrique

2.1 Chroniques et analyse non standard

L'*analyse non standard*, qui repose sur la logique mathématique, fournit un sens précis aux notions d'infiniment petit et d'infiniment grand. On recommande la présentation par Nelson (1977) (voir aussi Nelson (1987), Diener et Diener (1995); Diener et Reeb (1989)), conceptuellement plus simple.

Soit l'intervalle $[0, 1]$, avec échantillonnage infinitésimal, comme souvent en analyse non standard,

$$\mathfrak{T} = \{0 = t_0 < t_1 < \dots < t_\nu = 1\}$$

où $t_{i+1} - t_i$, $0 \leq i < \nu$, est *infinitésimal*, c'est-à-dire très petit.

Remarque 2.1 *Les notions d'infiniment petit ou grand sont des idéalizations mathématiques. On doit considérer, en pratique, un laps de temps d'une seconde (resp. heure) comme très petit par rapport à une heure (resp. mois). C'est pourquoi l'analyse non standard s'applique au monde réel.*

Une chronique X est une fonction $\mathfrak{T} \rightarrow \mathbb{R}$. La *mesure de Lebesgue* sur \mathfrak{T} est la fonction ℓ définie sur $\mathfrak{T} \setminus \{1\}$ par $\ell(t_i) = t_{i+1} - t_i$. La mesure de tout intervalle $[c, d] \subset \mathfrak{T}$, $c \leq d$, est sa longueur $d - c$. L'*intégrale* sur

$[c, d]$ de la chronique $X(t)$ s'écrit

$$\int_{[c,d]} X d\tau = \sum_{t \in [c,d]} X(t)\ell(t)$$

X est *S-intégrable* si, et seulement si, l'intégrale $\int_{[c,d]} |X| d\tau$ est *limitée*, c'est-à-dire non infiniment grande, et, si $d - c$ is infinitésimal, $\int_{[c,d]} |X| d\tau$ l'est aussi. X est *S-continue* en $t_\nu \in \mathfrak{T}$ si, et seulement si, $f(t_\nu) \simeq f(\tau)$ quand $t_\nu \simeq \tau$ (on écrit $a \simeq b$ si $a - b$ est infinitésimal). X est *presque continue* si, et seulement si, elle est *S-continue* sur $\mathfrak{T} \setminus R$, où l'ensemble R est *rare*¹. X est *Lebesgue intégrable* si, et seulement si, elle est *S-intégrable* et *presque continue*.

La chronique X est *rapidement fluctuante*, ou *oscillante*, autour de 0, si, et seulement si, elle est

- *S-intégrable*,
- $\int_A X d\tau$ est infinitésimal pour ensemble *quadrable*².

Théorème de Cartier-Perrin : Si la chronique X est *S-intégrable*, il y a décomposition additive³ (Cartier et Perrin (1995)) :

$$\boxed{X(t) = E(X)(t) + X_{\text{fluctuat}}(t)} \quad (1)$$

où

- la *moyenne* $E(X)(t)$ est Lebesgue-intégrable,
- $X_{\text{fluctuat}}(t)$ est *rapidement fluctuante*.

La décomposition (1) est unique à une quantité additive infinitésimale près.

Remarque 2.2 *Cette moyenne, ou tendance, souvent désignée par le mot anglais « trend », est à rapprocher du lissage par moyenne glissante ou « moving average » (voir, par exemple, Mélard (2008)). On la retrouve aussi en analyse technique (voir, par exemple, Béchu et al. (2014); Kirkpatrick et Dahlquist (2016); Tsinaslanidis et Zapranis (2016)), branche de l'ingénierie financière. Très différent est le sens habituel de trend dans la littérature sur les chroniques (voir, par exemple, Enders (2014)).*

2.2 Estimation algébrique des dérivées

Le § 2.1 démontre que l'on peut passer au continu si les données sont en nombre suffisant et la période d'échantillonnage petite. La dérivée est, en fait, estimée sur la moyenne et non sur les données brutes.

Soit la fonction polynômiale de degré 1

$$p_1(t) = a_0 + a_1 t, \quad t \geq 0, \quad a_0, a_1 \in \mathbb{R}$$

1. R est *rare* (Cartier et Perrin (1995)) si, pour tout réel standard $\alpha > 0$, il existe un ensemble interne $A \supset R$ tel que $\ell(A) \leq \alpha$.

2. Un ensemble est *quadrable* (Cartier et Perrin (1995)) si sa frontière est *rare*.

3. Voir Lobry et Sari (2008) pour une présentation plus abordable

Le calcul opérationnel classique (voir, par exemple, Yosida (1984)) permet de réécrire p_1

$$P_1 = \frac{a_0}{s} + \frac{a_1}{s^2}$$

D'où en multipliant par s^2

$$s^2 P_1 = a_0 s + a_1 \quad (2)$$

et en dérivant par rapport à s , qui, dans le domaine temporel, correspond à la multiplication par $-t$:

$$s^2 \frac{dP_1}{ds} + 2sP_1 = a_0 \quad (3)$$

Le système triangulaire (2)-(3) d'équations fournit a_0 , a_1 . On se débarrasse de sP_1 , s^2P_1 , an $s^2 \frac{dP_1}{ds}$, c'est-à-dire, dans le domaine temporel, des dérivées par rapport au temps, en multipliant les deux membres des équations (2) et (3) par s^{-n} , $n \geq 2$. Les intégrales itérées correspondantes, qui, en pratique, sont remplacées par des filtres linéaires numériques, sont des filtres passe-bas : elles atténuent les oscillations parasites. Une fenêtre temporelle courte suffit pour estimer avec précision a_0 , a_1 .

La généralisation aux polynômes de degré supérieur est évidente, donc aussi celle aux développements de Taylor tronqués. Renvoyons à Mboup et al. (2009) pour plus de détails.

Remarque 2.3 *Les manipulations algébriques précédentes sont employées avec succès en ingénierie (voir, par exemple, Sira-Ramírez et al. (2013)). Le passage en 0 de la dérivée première a été utilisée, en particulier, en traitement du signal (Fedele et al. (2009)).*

3 Illustrations

3.1 Polynômes par morceaux

Soit $y(t)$ défini par

$$\begin{cases} y(t) = -10t + 0.50 & 0 \leq t < 10 \\ y(t) = 5t - 99.5 & 10 \leq t < 20 \\ y(t) = -49.5 & 20 \leq t < 25 \\ y(t) = t^2 - 49.5 & 25 \leq t < 35 \\ y(t) = 50.5 & 35 \leq t \leq 45 \end{cases}$$

altéré par un bruit uniformément réparti entre -5 et 5 . Soit $0.1s$ la période d'échantillonnage. La segmentation obtenue est présentée figure 2-(a). L'estimation de dérivée l'est, selon le § 2.2, avec un polynôme de degré 2 sur une fenêtre de 40 échantillons. La figure 2-(b) en atteste la qualité. On note la détection des plateaux. La segmentation est peu représentative durant la branche parabolique, de courte durée heureusement⁴.

4. Voir, à ce sujet, le point 1 du § 4.

3.2 Sinusoïdes

Soient deux sinusoïdes de fréquences 5s et 30s, échantillonnées toutes les 0.01s (voir figures 3-(a) et 4-(a)). Elles sont altérées par un bruit uniformément réparti entre -0.25 et 0.25 . La dérivée est estimée avec un polynôme de degré 2, sur une fenêtre glissante de durée 2s. On détecte les extrema avec un seuil égal à 0.2 (voir figures 3-(b) et 4-(b)). Avec la petite fréquence, ce seuillage conduit à un plateau (figures 4-(c) & 4-(d)), mais pas avec la grande (figures 3-(c) & 3-(d)).

3.3 CAC 40

On reprend, pour la segmenter, la chronique du CAC 40, représentée par la figure 1. Avec des seuils différents, 1 et 5, pour détecter les passages par zéro de la dérivée, on modifie notablement les résultats (figures 5-(c) et 6-(c)).

4 Conclusion

1. On examinera d'autres points d'importance perceptuelle, comme ceux de courbure localement maximale (voir, déjà, Join et Tabbone (2008)).
2. Aux chroniques multivariées correspondent des courbes gauches, c'est-à-dire non planes. Le choix des PIP exige une étude géométrique plus délicate (voir, par exemple, Cagnac et al. (1971)). L'analyse technique, qui examine, souvent, prix et volumes (Béchu et al. (2014); Kirkpatrick et Dahlquist (2016); Tsinaslanidis et Zapranis (2016)), y gagnerait certainement⁵.
3. Des applications véritables sont en cours d'étude, la reconnaissance de signatures entre autres. Plusieurs publications proposent, déjà, des techniques diverses d'analyse des chroniques (voir, par exemple, Henniger et al. (2014); Hsu et al. (2015)). Nos premiers résultats démontent l'inanité de rechercher mesures ou distances universelles de similarité, c'est-à-dire indépendantes du but poursuivi.

5. Certains liens entre analyse technique, PIP et segmentation des chroniques ont déjà été entrevus (voir, par exemple, Tsinaslanidis et Zapranis (2016); Wan et al. (2016); Wan et Si (2017); Yin et al. (2011)).

Références

- Abonyi, J., Feil, B. (2007) : Cluster Analysis for Data Mining and System Identification. Birkhäuser.
- Abouaïssa, H., Fliess, M., Join, C. (2016) : On short-term traffic flow forecasting and its reliability. In: 8th IFAC Conf. Manufac. Model. Manag. Contr., Troyes. <https://hal.archives-ouvertes.fr/hal-01275311/en/>
- Béchu, T., Bertrand, E. Nebenzahl, J. (2014) : L'analyse technique (7^e éd.). Economica.
- Cagnac, G., Ramis, E., Commeau, J. (1971) : Traité de mathématiques spéciales, t. 4. Masson.
- Cartier, P., Perrin, Y. (1995) : Integration over finite sets. In: Diener, F., Diener, M. (eds.) Nonstandard Analysis in Practice, pp. 195-204. Springer.
- Castellini, A., Bicego, M., Masillo, F., Zuccotto, M., Farinelli, A. (2020) : Time series segmentation for state-model generation of autonomous aquatic drones : A systematic framework. Engin. Appli. Artif. Intell. **90**, 103499.
- Cheong, S.A., Forna, R.P., Hui Ting Lee, G., Liang Kok, J., Shyr Yim, W., Yuan Xu, D., Zhang, Y. (2012) : The Japanese economy in crises: A time series segmentation study. Econom.: The Open-Access Open-Assess. E-J. **6**. <http://dx.doi.org/10.5018/economics-ejournal.ja.2012-5>
- Cho, H., Fryzlewicz, P. (2012) : Multiscale and multi-level technique for consistent segmentation of nonstationary time series. Stat. Sinica **22**, 207-229.
- Diener, F., Diener, M. (1995) : Tutorial. In: Diener, F., Diener, M. (eds.) Nonstandard Analysis in Practice, pp. 1-21. Springer.
- Diener, F., Reeb, G. (1989) : Analyse non standard. Hermann.
- Durán-Rosal, A.M., Gutiérrez, P.A., Martínez-Estudillo, F.J., Hervas-Martínez, C. (2018a) : Simultaneous optimisation of clustering quality and approximation error for time series segmentation. Informat. Sci. **442-443**, 186-201.
- Durán-Rosal, A.M., Gutiérrez, P.A., Salcedo-Sanz, S., Hervas-Martínez, C. (2018b) : A statistically-driven Coral Reef Optimization algorithm for optimal size reduction of time series. Appl. Soft Comput. **63**, 139-153.
- Durán-Rosal, A.M., Gutiérrez, P.A., Carmona-Poyato, Á, César Hervas-Martínez, C. (2019) : A hybrid dynamic exploitation barebones particle swarm optimisation algorithm for time series segmentation. Neurocomput. **353**, 45-55.
- Edelhoff, H., Signer, J., Balkenhol, N. (2016) : Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. Mov. Ecol. **4**, 21. <https://doi.org/10.1186/s40462-016-0086-5>
- Esling, P., Agon, C. (2012) : Time-series data mining. ACM Comput. Surv. **45**, article 12, 34 pages.
- Enders, W. (2014) : Applied Econometric Time Series (4th ed.). Wiley.
- Fedele, G., Chiaravallotti, F., Join, C. (2009) : An algebraic derivative-based approach for the zero-crossings estimation. In: 17th Europ. Sign. Proc. Conf., Glasgow. <https://hal.inria.fr/inria-00389872/en/>
- Fink, E., Pratt, K.B. (2004) : Indexing of compressed time series. In: Last, M., Kandel, A., Bunke, H. (eds) Data Mining in Time Series Databases, pp. 43-65. World Scientific.
- Fink, E., Gandhi, H.S. (2011) : Compression of time series by extracting major extrema. J. Exper. Theoret. Artif. Intell. **23**, 255-270.
- Fliess, M. (2006) : Analyse non standard du bruit. C.R. Acad. Sci. Paris **342**, 797-802.
- Fliess, M., Join, C. (2009) : A mathematical proof of the existence of trends in financial time series. In: El Jai, A., Afifi, L., Zerrik, E. (Eds.): Systems Theory: Modeling, Analysis and Control. Presses Universitaires de Perpignan, pp. 43-62. <https://hal.inria.fr/inria-00352834/en/>
- Fliess, M., Join, C., Bekcheva, M., Moradi, A., Mounier, H. (2019) : Easily implementable time series forecasting techniques for resource provisioning in cloud computing. In: 6th Int. Conf. Contr. Dec. Inform. Techno., Paris. <https://hal.archives-ouvertes.fr/hal-02024835/en/>
- Fliess, M., Join, C., Hatt, F. (2011) : A-t-on vraiment besoin d'un modèle probabiliste en ingénierie financière? In: Conf. Médit. Ingén. Sûre Syst. Compl., Agadir. <https://hal.archives-ouvertes.fr/hal-00585152/en/>
- Fliess, M., Join, C., Mboup, M. (2010) : Algebraic change-point detection. App. Algeb. Engin. Commun. Comput. **21**, 131-143.
- Fliess, M., Join, C., Sira-Ramírez, H. (2008) : Non-linear estimation is easy. Int. J. Model. Identif. Contr. **4**, 12-27.
- Fliess, M., Join, C., Voyant, C. (2018) : Prediction bands for solar energy: New short-term time series forecasting techniques. Solar Energy **166**, 519-528.
- Fu, T.-C. (2011) : A review on time series data mining. Engin. Appli. Artif. Intell. **24**, 164-181.

- Fu, T.-C., Hung, Y.-K., Chung, F.-L. (2017) : Improvement algorithms of perceptually important point identification for time series data mining. In: IEEE 4th Int. Conf. Soft Comput. Mach. Intell.. Port Louis, Mauritius.
- Henniger, O., Guest, R., Miguel-Hurtado, O., Kaplan, C. (2014) : Signature/Sign Time Series Data: Standardization. In: Li S., Jain A. (Eds), Encyclopedia of Biometrics, pp. 1395-1400. Springer.
- Hsu, Y.-L., Chu, C.-L., Tsai, Y.-J., Wang, J.-S. (2015) : An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition. IEEE Sensors J. **15**, 154-163.
- Hubert, P., Carbonnel, J.P., Chaouche, A. (1989) : Segmentation des séries hydrométéorologiques – Applications à des séries de précipitations et de débit de l’Afrique de l’ouest. J. Hydrol. **110**, 349-367.
- Join, C., Tabbone, S. (2008) : Robust curvature extrema detection based on new numerical derivation. In: Blanc-Talon, J., Bourennane, S., Philips, W., Popescu, D., Scheunders, P. (Eds): Advanced Concepts for Intelligent Vision Systems. Lect. Notes Comput. Sci. **5259**, pp. 485-493. Springer.
- Keogh, E., Chu, S., Hart, D., Pazzani, M. (2004) : Segmenting time series: A survey and novel approach. In : Last, M., Kandel, A., Bunke, H. (eds), Data Mining in Time Series Databases, pp. 1-21. World Scientific.
- Kim, H., Kim, H.K., Kim, M., Park, J. , Cho, S., Im, K.B., Ryu, C.R. (2019) : Representation learning for unsupervised heterogeneous multivariate time series segmentation and its application. Comput. Indust. Engin. **130**, 272-281
- Kirkpatrick, C.D., Dahlquist, J. (2016) : Technical Analysis (3rd ed.). Pearson.
- Lanczos, C. (1956) : Applied Analysis. Prentice Hall.
- Lee, W.-H., Ortiz, J., Ko, B., Lee, R. (2018) : Time series segmentation through automatic feature learning. [arXiv:180.955394](https://arxiv.org/abs/180.955394)
- Lobry, C., Sari, T. (2008) : Nonstandard analysis and representation of reality. Int. J. Contr. **39**, 535-576.
- Liu, X., Lin, Z., Wang, H. (2008) : Novel online methods for time series segmentation. IEEE Trans. Knowledg. Data Engin., **20**, 1616-1626.
- Liu, J., Wang, C., Liu, Y. (2019) : A novel method for temporal action localization and recognition in untrimmed video based on time series segmentation. IEEE Access **7**, 135204-135209.
- Lu, S., Huang, S. (2020) : Segmentation of multivariate industrial time series data based on dynamic latent variable predictability. IEEE Access **8**, 112092-112103.
- Martí, L., Sanchez-Pi, N., Molina, J.M., Bicharra García, A.C. (2014) : YASA: Yet another time series segmentation algorithm for anomaly detection in big data problems. In: Polycarpou, M., de Carvalho, A.C.P.L.F., Pan, J.-S., Woźniak, M., Quintián, H., Corchado, E. (Eds.): Hybrid Artificial Intelligence Systems. Lect. Notes Artif. Intel. **8480**, pp. 697-708. Springer.
- Mboup, M., Join, C., Fliess, M. (2009) : Numerical differentiation with annihilators in noisy environment. Numer. Algor. **50**, 439-467.
- Mélard, G. (2008) : Méthodes de prévision à court terme (2^e éd.). Ellipses & Presses Universitaires de Bruxelles.
- Nelson, E. (1977) : Internal set theory: A new approach to nonstandard analysis. Bull. Amer. Math. Soc. **83**, 1165-1198.
- Nelson, E. (1987) : Radically Elementary Probability Theory. Princeton University Press.
- Pratt, K.B., Fink, E. (2002) : Search for patterns in compressed time series. Int. J. Image Graph. **2**, 89-106.
- Robinson, A. (1996) : Non-standard Analysis (revised ed.). Princeton University Press.
- SeEVERS, J.-P., Johst, J., Weiß, T., Meschede, H., Hessel, J. (2019) : Automatic time series segmentation as the basis for unsupervised, non-intrusive load monitoring of machine tools. Procedia CIRP **81**, 695-700.
- Sira-Ramírez, H., García-Rodríguez, C., Cortès-Romero, J., Luviano-Juárez, A. (2013) : Algebraic Identification and Estimation Methods in Feedback Control Systems. Wiley.
- Tsinaslanidis, P.E., Zapranis, A.D. (2016) : Technical Analysis for Algorithmic Pattern Recognition. Springer.
- Wan, Y., Gong, X., Si, Y.-W. (2016) : Effect of segmentation on financial time series pattern matching. Appl. Soft Comput. **38**, 346-359.
- Wan, Y., Si, Y.W. (2017) : A formal approach to chart patterns classification in financial time series. Informat. Sci. **411**, 151-175.
- Yin, J., Si, Y.-W., Gong, Z. (2011) : Financial time series segmentation based on turning points. In: Int. Conf. Syst. Sci. Engin., Macau.
- Yosida, K. (1984) : Operational Calculus (translated from the Japanese). Springer.

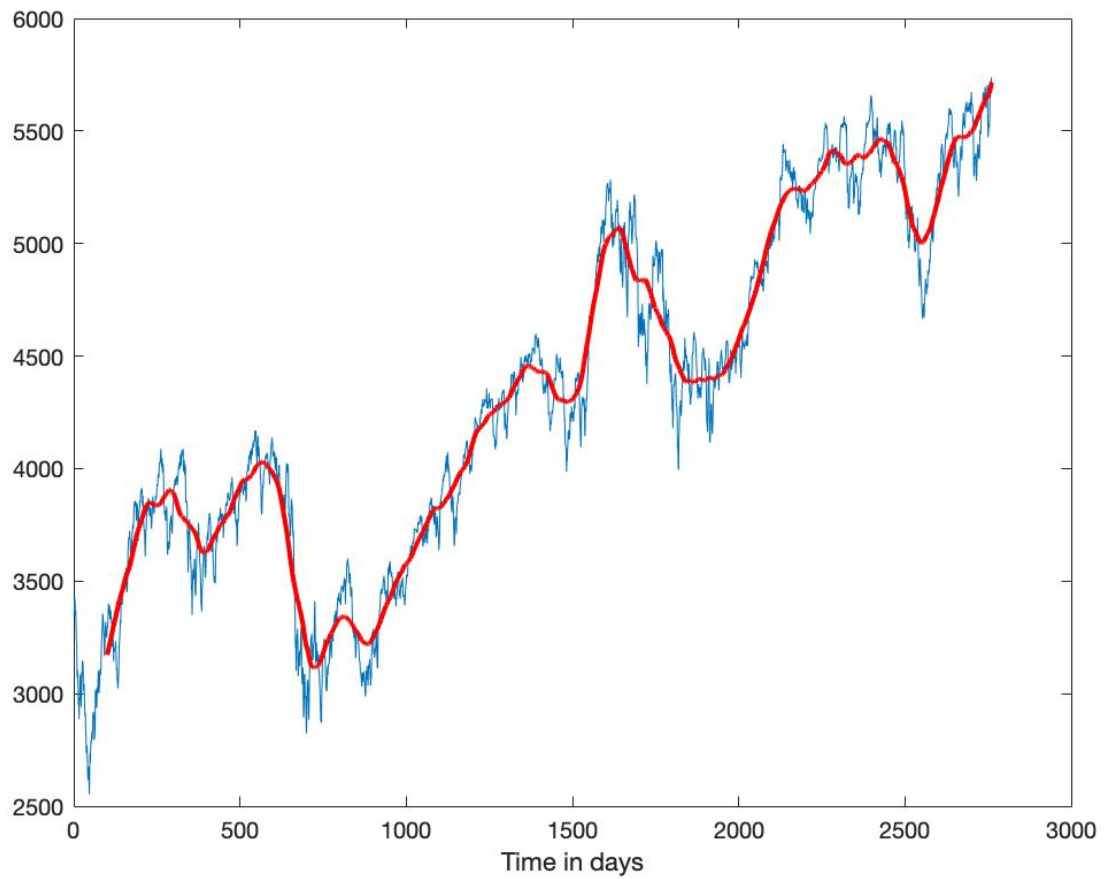
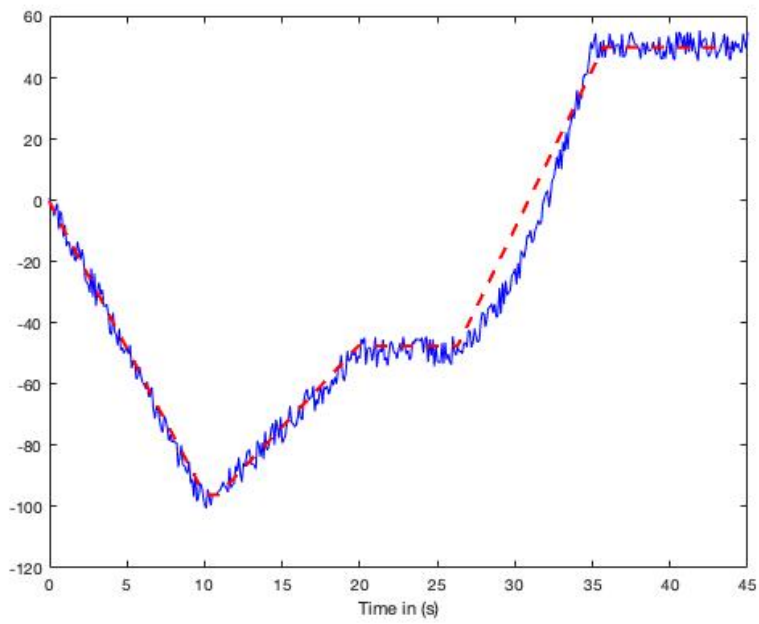
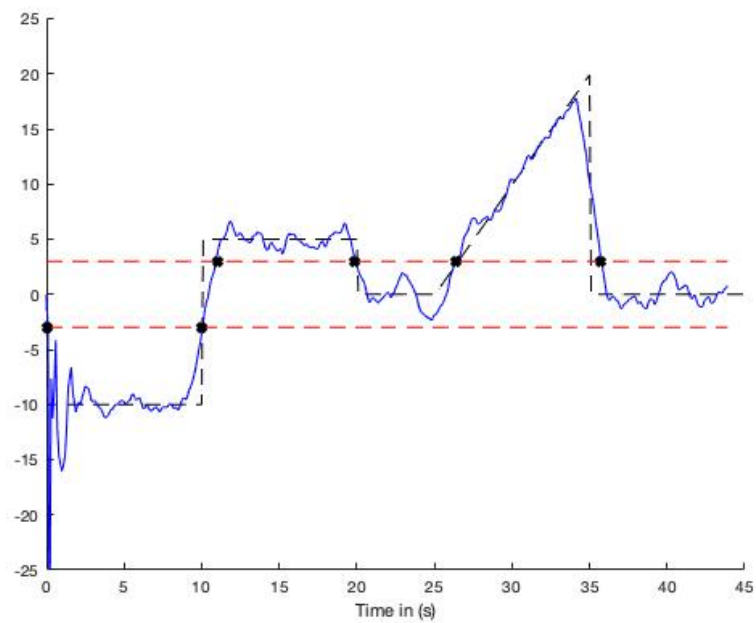


FIGURE 1 – CAC 40. En bleu : valeur quotidienne du CAC 40 du 1er janvier 2009 au 31 décembre 2019. En rouge : moyenne (glissante).

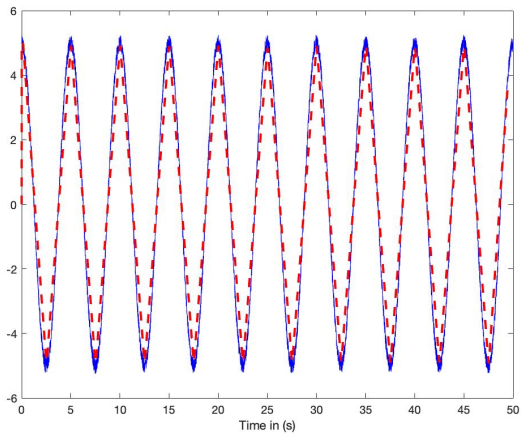


(a) En bleu : signal. En rouge : segmentation

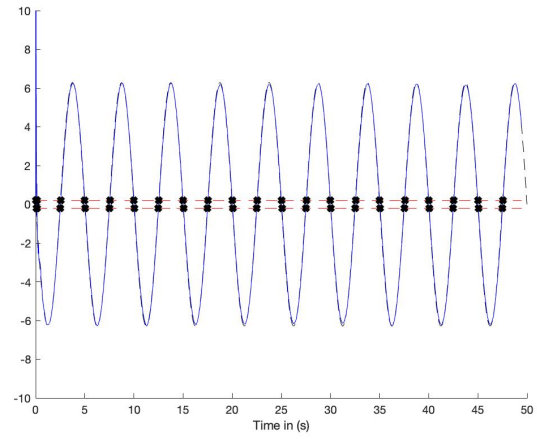


(b) Véritable dérivée (noir - -), dérivée estimée (bleu), seuils (rouge - -), passage en 0 (noir *)

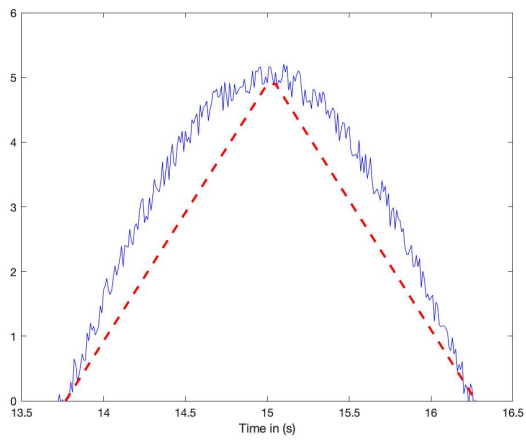
FIGURE 2 – Polynômes par morceaux



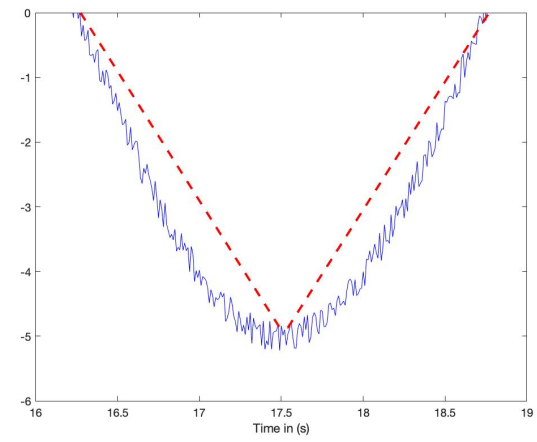
(a) En bleu : signal. En rouge : segmentation



(b) Véritable dérivée (noir - -), dérivée estimée (bleu), seuil (rouge - -), passage en 0 (noir *)

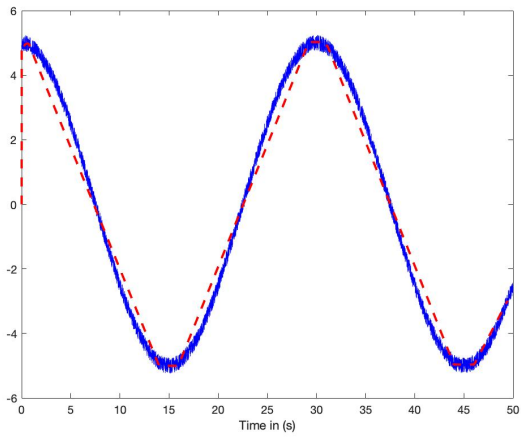


(c) Zoom de (a)

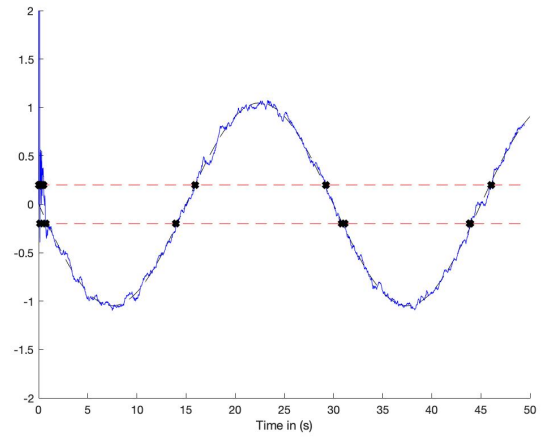


(d) Zoom de (a)

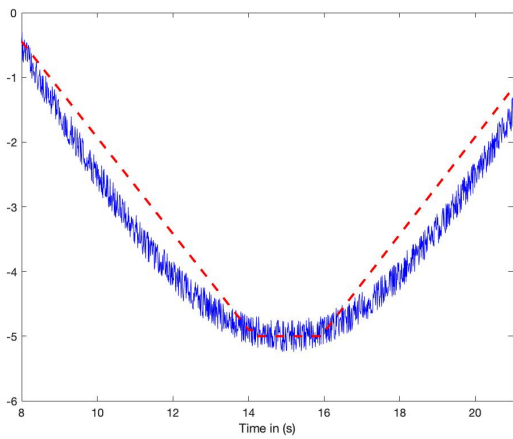
FIGURE 3 – Sinusoïde : fréquence élevée



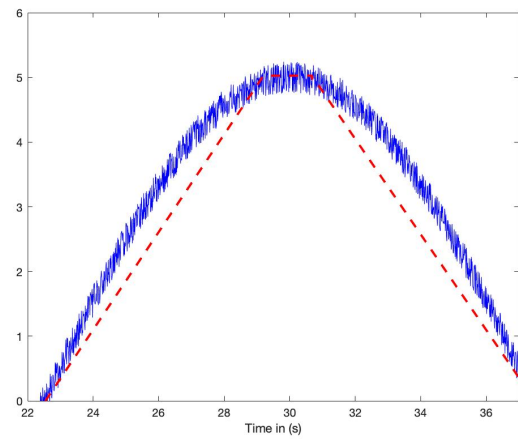
(a) En bleu : signal. En rouge : segmentation



(b) Véritable dérivée (noir - -), dérivée estimée (bleu), seuil (rouge - -), passage en 0 (noir *)

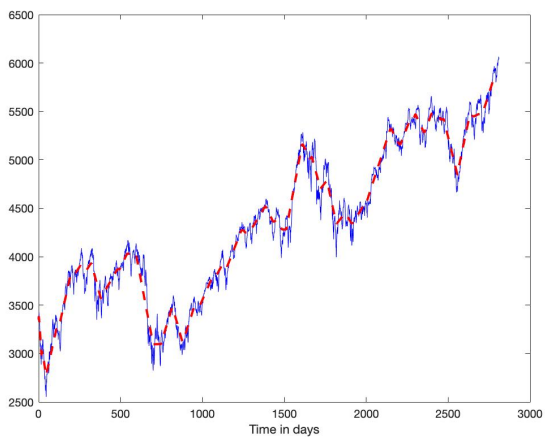


(c) Zoom de (a)

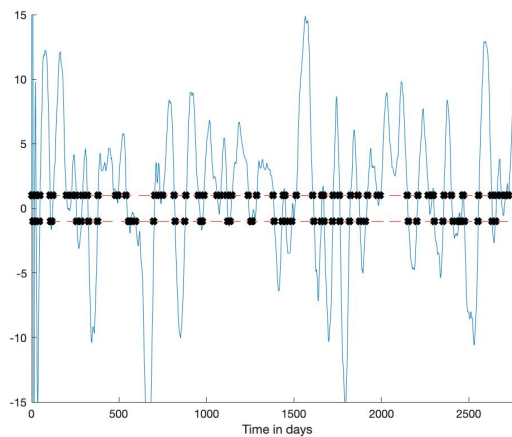


(d) Zoom de (a)

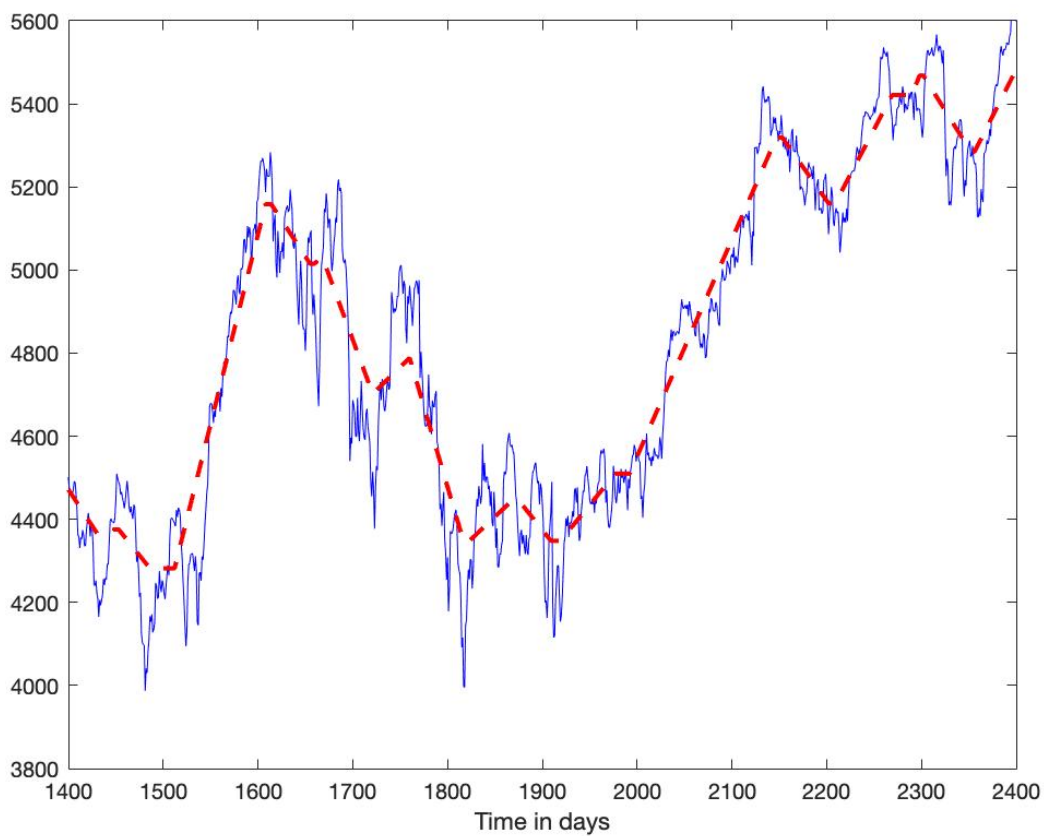
FIGURE 4 – Sinusoïde : fréquence basse



(a) En bleu : chronique. En rouge : segmentation

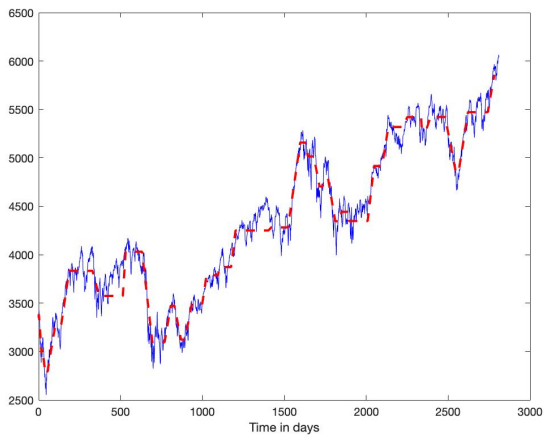


(b) Dérivée estimée (bleu), seuil (rouge - -), passage en 0 (noir *)

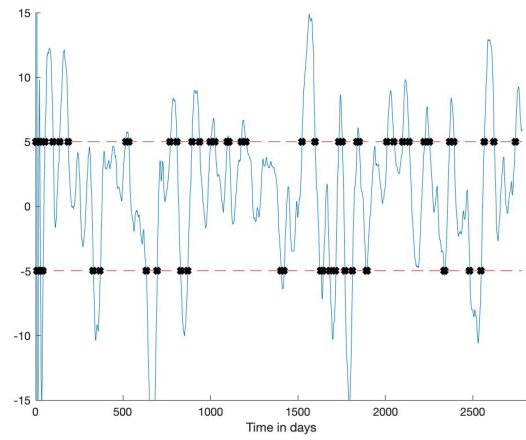


(c) Zoom de (a)

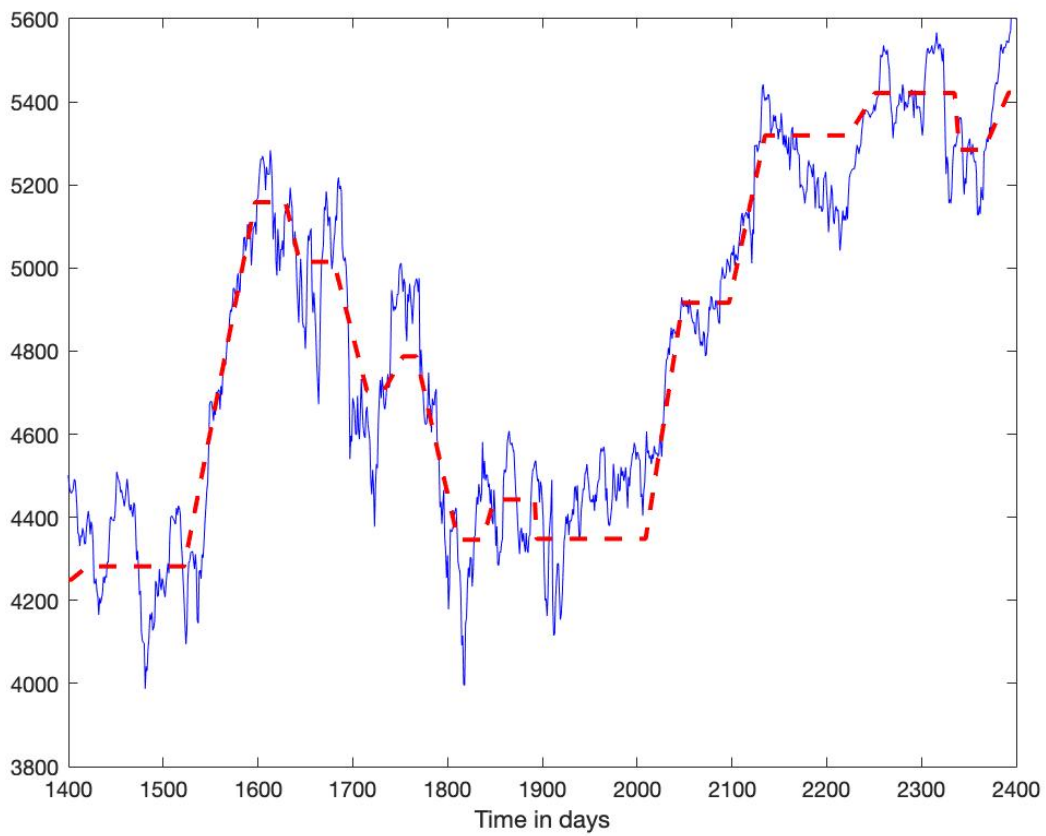
FIGURE 5 – CAC 40 : Seuil bas



(a) En bleu : chronique. En rouge : segmentation



(b) Dérivée estimée (bleu), seuil (rouge - -), passage en 0 (noir *)



(c) Zoom de (a)

FIGURE 6 – CAC 40 : Seuil haut