p-leader Multifractal Analysis and Sparse SVM for Intrapartum Fetal Acidosis Detection

R. Leonarduzzi[1], J. Spilka[2], J. Frecon[1], H. Wendt[3], N. Pustelnik[1], S. Jaffard[4], P. Abry[1], M. Doret[5],

Abstract— Interpretation and analysis of intrapartum fetal heart rate, enabling early detection of fetal acidosis, remains a challenging signal processing task. Among the many strategies that were used to tackle this problem, scale-invariance and multifractal analysis stand out. Recently, a new and promising variant of multifractal analysis, based on *p*-leaders, has been proposed. In this contribution, we use sparse support vector machines applied to *p*-leader multifractal features with a double aim: Assessment of the features actually contributing to classification; Assessment of the contribution of non linear features (as opposed to linear ones) to classification performance. We observe and interpret that the classification rate improves when small values of the tunable parameter *p* are used.

I. INTRODUCTION

Fetal monitoring. Fetal monitoring is commonly performed using cardiotocography, the simultaneous recording of fetal heart rate (FHR) and uterine contractions. It permits clinicians to identify fetuses at risk before asphyxia occurs, potentially inducing severe long term consequences, such as neuro-development disability, neonatal encephalopathy, and cerebral palsy. Continuous FHR offers valuable information about the fetal oxygenation status, and provides insight into the defense mechanisms fetuses use to adapt to hypoxia [1]. Fetuses reactions to hypoxic events result in a complex behaviour governed by multiple chemical and nervous feedback loops that cause complex FHR dynamics. In clinical practice, FHR is analyzed by considering short/long term variability and shifts in baseline FHR and deceleration shapes [2].

Scale invariance and multifractal analysis. Instead of performing the analysis of variability at pre-defined time scales, as proposed e.g., in the FIGO clinical guidelines [2], the scale invariance paradigm measures the *relationship* between the data variability across several scales $a \in [a_m, a_M]$. Variability is measured by the q-th order sample moment $S(q, a)$ of a well-chosen *multiresolution quantity* T_X. When data show some form of scale invariance, S follows a power-law behavior across the scales: $S(q, a) \sim a^{qH}$, $a \to 0^+$, with scaling exponent qH, controlled by the *Hurst parameter H*.

Multifractal analysis extends the characterization provided by self-similar scale-invariant models by allowing the power-law of S to be controlled by a nonlinear function of q: $S(q, a) \sim a^{\zeta(q)}$. $\zeta(q)$ is called *scaling function*. Scale invariance and multifractal paradigms have already been proven useful for the analysis of heart rate data, both in adults and during labour, cf. e.g. [3], [4].

p-leader based multifractal analysis. Classical multiresolution quantities for scale invariance and multifractal analysis are wavelet coefficients and wavelet leaders, cf., e.g., [5]. However, it has recently been proposed to base multifractal estimations on *p-leaders* [6], which provide a number of benefits over the traditional wavelet leaders. First, they allow to estimate a scaling function $\zeta(q)$ that is negative for large q; this behavior is related to the presence of negative regularity, typically present in heart rate data [4], [7]. Second, the potential dependence on the parameter p provides refined and detailed analysis in the regularity of FHR fluctuations along time. Finally, estimations based on *p*-leaders show improved statistical performance compared to others based on more classical quantities [6].

Goals and contributions. *p*-leader multifractal analysis has already been explored for acidosis detection in intrapartum FHR in [4], yet leaving two issues unaddressed: i) to what extent does the information encoded in the scaling function (a nonlinear feature) improves acidosis detection compared to using only the Hurst parameter (a linear feature) ii) Does acidosis detection performance depend on parameter p? To address these issues, we propose here to make use of sparse support vector machines [8], a method recently introduced for supervised classification that aims to achieve jointly both feature selection and efficient classification.

Outline. The large size database used here is described in Sec. II. Brief introductions to *p*-leader multifractal analysis and Sparse-SVM are given in Sec. III. Results and conclusions are discussed in Sec. IV and Sec. V, respectively.

II. DATABASE

Data. The large database of FHR signals was collected in routine clinical practice at the French public academic Hospital Femme-Mère-Enfant in Bron, between 2000 and 2010. In total, the database consists of 3049 intrapartum cardiotocogram (CTG) signals – all acquired using a scalp electrode system STAN S21 or S31, with 12-bit resolution, 500 Hz sampling frequency (STAN, Neoventa Medical, Moelndal, Sweden). Clinical information for women and neonates was systematically collected by obstetricians in charge of the delivery, cf. [9] for details.

[1] Physics Dept., ENS Lyon, CNRS, France `roberto.leonarduzzi@ens-lyon.fr`, `patrice.abry@ens-lyon.fr`, `jordan.frecon@ens-lyon.fr`, `nelly.pustelnik@ens-lyon.fr`
[2] Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, CZ, `jiri.spilka@ciirc.cvut.cz`
[3] IRIT-ENSEEIHT, Université de Toulouse, UMR, France `herwig.wendt@irit.fr`
[4] LAMA, Math. Dept., Université Paris-Est, Créteil, France `jaffard@u-pec.fr`
[5] Hospices Civil de Lyon, Femme-Mère-Enfant Hospital, Lyon, France `muriel.doret@chu-lyon.fr`

Dilation Stage. In regular delivery, two stages of labour are identified: the dilation and the active pushing stages. Because the temporal dynamics of both stages are different [10], focus is here on the end of the dilation stage. Thus, subjects that fulfilled the following criteria were used: i) had active pushing stage shorter than 20 minutes, or underwent a c-section, and ii) had umbilical artery pH measurement immediately after birth. This leads to selecting 1288 records, amongst which 37 fetuses have neonatal acidosis (umbilical artery pH ≤ 7.05). Hereafter, we refer to this latter group as *nonhealthy*, as opposed to *healthy*. The last 25 minutes of FHR at the end of the dilation stage are analyzed (disregarding the very last 5 minutes for data quality issues).

Preprocessing. FHR data consists of time intervals between consecutive R peaks $\{t_i, i = 1, \ldots, N\}$. Classically, in FHR variability analysis, beat to beat time series are interpolated and resampled into regularly sampled signals, here at 10 Hz using cubic spline interpolation, to produce the beat-per-minute (bpm) signal $X(t)$.

III. METHODS

Wavelet coefficients. Let ψ denote the mother wavelet, characterized by its number of vanishing moments N_ψ, a strictly positive integer such that $\int_{\mathbb{R}} t^k \psi(t) dt = 0, \forall k = 0, \ldots, N_\psi - 1$, and $\int_{\mathbb{R}} t^{N_\psi} \psi(t) dt \neq 0$. Let $\{\psi_{j,k}(t) = 2^{-j}\psi(2^{-j}t - k)\}_{(j,k) \in \mathbb{N}^2}$ be the orthonormal basis of $L^2(\mathbb{R})$ formed by dilations and translations of ψ. The (L^1-normalized) discrete wavelet transform coefficients are defined as $c_{j,k} = 2^{-j} \langle \psi_{j,k} | X \rangle$ (cf. e.g. [11], for details on wavelet transforms).

Hurst exponent. Wavelet coefficients permit a simple, robust and efficient estimation of the Hurst parameter [3], [5]. The sample moment estimate of the variance of wavelet coefficients is computed: $S(j) = 1/n_j \sum_{k=1}^{n_j} c_{j,k}^2$ (with n_j the number of wavelet coefficients available at scale j). It behaves as $S(j) = K 2^{2jH}$ for self-similar processes. Therefore, H can be estimated by a simple linear regression of $\log_2 S(j)$ against j.

Wavelet p-leaders. The p-leaders are defined as local L^p-norms of scaled wavelet coefficients [6], [12]

$$\ell_{j,k}^{(p)} \triangleq \left(\sum_{\lambda' \subset 3\lambda} |2^{j'\gamma} c_{\lambda'}|^p \ 2^{j'-j} \right)^{\frac{1}{p}}, \qquad (1)$$

with $\lambda = \lambda_{j,k} = [k2^j, (k+1)2^j)$, $c_\lambda = c_{j,k}$ and $3\lambda = \bigcup_{m \in \{-1,0,1\}} \lambda_{j,k+m}$. That is, the local norm considers all wavelet coefficients in a narrow time neighbourhood of $t = 2^j k$, and for all finer scales $j' \leq j$. The parameter $\gamma \geq 0$ must be chosen to ensure a minimal regularity constraint (cf. [12] for details). The parameter p can be freely chosen such that $p \in (0, p_0)$, where p_0 is implicitly defined by

$$\eta(p_0) + \gamma p_0 = 0, \qquad (2)$$

where $1/n_j \sum_{k=1}^{n_j} |c_{j,k}|^p \sim 2^{j\eta(p)}, j \to 0$. The p-leaders allow to measure the p-exponent, which quantifies the local regularity fluctuations of X (cf. [12] for details, outside of the scope of this contribution). Also, note that classical wavelet leaders are given for $p = +\infty$, in which case (1) reduces to $\ell_{j,k}^\infty \triangleq \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|$.

Log-cumulants. Instead of computing the function $\zeta(q)$ directly from the mesurement of the q-th moments of the p-leaders, we will follow [13] and compute a polynomial approximation instead: $\zeta(q) = \sum_{m=1}^{\infty} c_m q^m / m!$. The coefficients c_m, which are called *log-cumulants*, are related to the decay across the scales of the cumulants $C_m^{(p)}(j) = \text{Cum}_m \ln \ell_{j,\cdot}^{(p)}$ of the random variables $\ln \ell_{j,k}^{(p)}$

$$C_m^{(p)}(j) = C_0^{(p)} + c_m^{(p)} \ln 2^j. \qquad (3)$$

Therefore, $c_m^{(p)}$ can be estimated by linear regressions of $C_m^{(p)}(j)$ against $\ln 2^j$. The coefficient c_1 measures the second order (or correlation) properties of X, whereas $c_m^{(p)}$, $m \geq 2$, characterize the higher order statistical behavior of X. For self-similar processes, ζ is a linear function: $\zeta(q) = qH$. Thus, $c_1^{(p)} = H$ and the non linear features are all equal to zero: $c_m^{(p)} = 0$, $m \geq 2$. In general, if the regularity condition (2) holds, then $c_1^{(p)}$ and H provide essentially the same information and are thus linear features. An example of the correlation coefficient between H and $c_m^{(p)}$ is given in Table I, for $p = 0.25$; it confirms that H and c_1 are strongly correlated. The same behavior is observed for other values of p. Thus, in the remainder of this work, only H and not c_1 is used. Coefficients $c_m^{(p)}$ are related to the shape of the *multifractal spectrum* of X, which provides a global description of the regularity fluctuations of X, they are hence non linear features. The interested reader is referred to e.g. [5], [12] for more details on multifractal analysis.

Sparse SVM. Support vector machines (SVM) search a decision function $D(\boldsymbol{x}) = \text{sgn}(\boldsymbol{w}^\top \boldsymbol{x} + b)$ which provides an optimal separation between two classes using a d-dimensional feature vector $\boldsymbol{x}$. To achieve jointly optimal classification and feature selection, it has been proposed to introduce sparsity into classical SVM formulations by use of L_1-norm soft regularization (cf. e.g., [8]). The resulting Sparse-SVM minimization problem reads

$$(\widehat{\boldsymbol{w}}, \widehat{b}) \in \underset{\boldsymbol{w} \in \mathbb{R}^N, \, b \in \mathbb{R}}{\arg\min} \ C \sum_{i=1}^{N} \max(0, 1 - y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b))^2 + \|\boldsymbol{w}\|_1,$$
$$(4)$$

where $y_i \in \{+1, -1\}$ and the regularization constant C controls the tradeoff between sparsity and data fidelity, formulated with the square hinge loss function [8]: Decreasing C favors sparsity. However, the nonsmooth nature of the L_1-norm sparsity term significantly complicates the resolution of the optimization problem (4), which then requires the use of a Forward-Backward Splitting Algorithm involving proximity operators to handle the nonsmoothness [14].

IV. RESULTS

Setup. Daubechies' wavelet with $N_\psi = 3$ vanishing moments are used. Linear regressions for the estimation of H and $c_m^{(p)}$, $m \in \{2, 3, 4\}$, are performed in the same scaling range $j \in [6, 10]$, corresponding to time scales ranging from 6 to 102.4 s. Inspection of the database leads to choose $\gamma = 1$ to ensure that the minimum regularity condition in eq. (2) holds for all records. The following representative values of

TABLE I

FEATURE CORRELATIONS ($p = 0.25$).

	H	$c_1^{(0.25)}$	$c_2^{(0.25)}$	$c_3^{(0.25)}$	$c_4^{(0.25)}$
H	1.00	0.82	0.02	0.11	0.07
$c_1^{(0.25)}$	0.82	1.00	-0.23	0.02	0.16
$c_2^{(0.25)}$	0.02	-0.23	1.00	0.16	-0.47
$c_3^{(0.25)}$	0.11	0.02	0.16	1.00	0.18
$c_4^{(0.25)}$	0.07	0.16	-0.47	0.18	1.00

TABLE II

UNIVARIATE PERFORMANCE.

Feature	AUC	SP	SE	#TP	#FP
H	0.71	0.70	0.62	26	464
$c_2^{(0.25)}$	0.63	0.70	0.46	26	663
$c_3^{(0.25)}$	0.63	0.70	0.43	26	696
$c_4^{(0.25)}$	0.57	0.70	0.46	26	663
$c_2^{(1)}$	0.62	0.70	0.38	26	763
$c_3^{(1)}$	0.59	0.70	0.41	26	730
$c_4^{(1)}$	0.62	0.70	0.43	26	696
$c_2^{(2)}$	0.62	0.70	0.46	26	663
$c_3^{(2)}$	0.59	0.70	0.43	26	696
$c_4^{(2)}$	0.62	0.70	0.46	26	663
$c_2^{(\infty)}$	0.62	0.70	0.43	26	696
$c_3^{(\infty)}$	0.60	0.70	0.43	26	696
$c_4^{(\infty)}$	0.58	0.70	0.43	26	696

Fig. 1. **Area under the ROC curve** (AUC) as a function of the regularization parameter C, for several values of p (left).

Fig. 2. **Selected features** ($p = 0.25$). Absolute values of the weights associated to features $\left(H, c_2^{(0.25)}, c_3^{(0.25)}, c_4^{(0.25)}\right)$ as functions of the regularization constant C.

p in p-leader analysis are chosen: $p \in \{0.25, 1, 2, 4, \infty\}$. Training and classification performance measurements are done using a 37-fold stratified cross-validation scheme.

Performance metrics. *Receiver operating characteristic* (ROC) curves [15] are constructed by shifting the classifier threshold computing the performance in each case. Then, the area under the ROC curve (AUC) is computed, with a simple numeric integral, as an indicator of classifier quality. Further, an operation point on the ROC curve is selected for a fixed *sensitivity* (SE) of 0.70, corresponding to the correct detection (TP) of 26 nonhealthy fetuses. The corresponding *specificity* (SP) and number of false positives (FP) are computed.

Univariate classification. As a first step, we perform univariate classification for each feature independently, using SVM classifier with univariate feature vector input. Table II reports the AUC, SP, SE and number of TP and FP. It shows that (linear) feature H outperforms all (nonlinear) multifractal features $c_m^{(p)}$: with AUC and SE larger by 0.1, and the lowest number of FP for TP= 26. Moreover, when compared across the different values of p, the univariate performance of $c_m^{(p)}$ are observed to be roughly equivalent.

Multivariate classification. Table I shows the correlation coefficient between features $H, c_1^{(p)}, c_2^{(p)}, c_3^{(p)}, c_4^{(p)}$ for $p = 0.25$ (as an example). It indicates weak correlations amongst used features, with the slight exception of c_2 and c_4 being weakly (anti-)correlated. Identical conclusions hold for all values of p. This suggests that multivariate classification can be successful. To evaluate the benefit of multivariate classificatio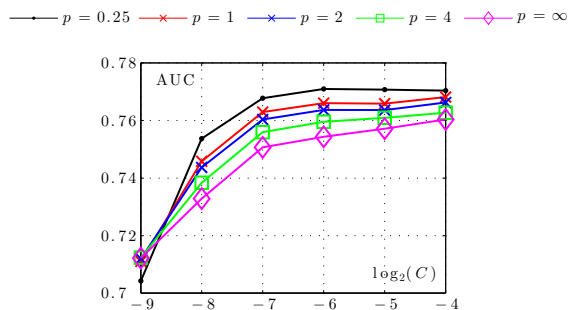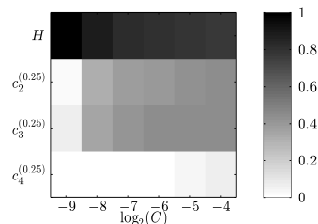n, with the combined use of linear and non linear features, Sparse-SVM is used with feature vectors of the form $\boldsymbol{x} = \left(H, c_2^{(p)}, c_3^{(p)}, c_4^{(p)}\right)$, for fixed p, as inputs.

Impact of p. Fig. 1 shows the classification performance measured by AUC as a function of the regularization constant C, for the five different values of p. It shows clearly that, irrespective of C, the AUC increases when p is decreased. This is in agreement with results in [6], reporting that p-leader based estimates yield improved classification performance in FHR compared to classical leader based ones ($p = \infty$). From now on, we therefore focus the analysis on performance obtained with $\boldsymbol{x} = \left(H, c_2^{(0.25)}, c_3^{(0.25)}, c_4^{(0.25)}\right)$. Note that linear feature H does not depend on p.

Feature selection. Fig. 2 reports the (absolute) values of the weights quantifying the contribution to classification of features $\left(H, c_2^{(0.25)}, c_3^{(0.25)}, c_4^{(0.25)}\right)$, as functions of the regularization constant C, for $p = 0.25$. For the lowest C, Sparse-SVM maximizes sparsity (as expected) and selects mostly linear feature H, ignoring nonlinear features $c_2^{(0.25)}, c_3^{(0.25)}, c_4^{(0.25)}$. When C is increased, features $c_2^{(0.25)}$ and $c_3^{(0.25)}$ are included, yet with smaller weights than that of H. Both features show similar weights and are thus granted similar importance by Sparse-SVM. This is consistent with univariate classification: The most discriminant univariate feature H is selected first. When sparsity is relaxed, c_2 and c_3, the second most discriminant univariate features, are involved. Feature c_4, with lowest univariate performance, is barely selected in multivariate classification. Equivalent conclusions are drawn when using any other p.

Linear vs. nonlinear features. Figs. 1, 2 and 3 combined

TABLE III

SPARSE-SVM PERFORMANCE ($p = 0.25$)

$\log_2(C)$	AUC	SP	SE	#TP	#FP
−9	0.70	0.70	0.59	26	497
−8	0.75	0.70	0.68	26	398
−7	0.77	0.70	0.68	26	398
−6	0.77	0.70	0.70	26	365
−5	0.77	0.70	0.70	26	365
−4	0.77	0.70	0.70	26	365

TABLE IV

MULTIFRACTAL ATTRIBUTES: MEAN (STANDARD DEVIATION) AND

P-VALUE OF WILCOXON RANK SUM TEST.

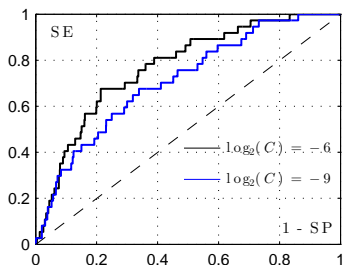|	Healthy	Nonhealthy	p-value
H	1.50 (0.481)	1.87 (0.423)	9.27E-6
$c_2^{(0.25)}$	-0.08 (0.051)	-0.10 (0.053)	7.19E-3
$c_3^{(0.25)}$	-1.50E-2 (3.50E-2)	2.03E-3 (2.80E-2)	1.47E-2
$c_4^{(0.25)}$	6.17E-3 (3.91E-2)	9.30E-3 (3.43E-2)	1.65E-2

Fig. 3. **ROC curves** for Sparse-SVM ($p = 0.25$).

show that optimal performance is achieved for a level of sparsity quantified by $\log_2 C = -6$. This corresponds to the combined use of features H, $c_2^{(0.25)}$ and $c_3^{(0.25)}$ and hence clearly shows that: i) Multivariate classification outperforms any univariate classifications; ii) The linear feature H is the one contributing most to classification, but nonlinear features do contribute to classification and improve performance. Fig. 3 notably shows that nonlinear features improve the entire ROC curve shape and thus contribute at all levels of specificity and sensitivity. For the lowest C, Sparse-SVM selects mostly H and the performance is equivalent for all p. When sparsity decreases, more weight is given to c_2 and c_3 and the AUC increases significantly and reaches a plateau around $\log_2 C = -6$. The same behavior is observed for all values of p. Table III reports the improvement in sensitivity obtained for a fixed specificity of 0.70. Comparing the number of FP in the first row (which is equivalent to the use of only H and hence to the first row in Table II) against the four last rows (which correspond to the use of H, c_2 and c_3) clearly shows that the incorporation of the nonlinear features permits a decrease of more than 100 FP. Linear parameter H essentially quantifies the distribution of energy across frequencies in the Fourier spectrum (thus the correlation or second order statistics). H is observed to be larger for nonhealthy subjects (cf. Table IV) corresponding to less power at high frequency and hence to weaker variability in the temporal frequencies of intrapartum FHR. Non linear parameters c_2 and c_3 quantify the dependence structure beyond correlation: Large values of $|c_2|$ indicate a significant burstiness in temporal dynamics; c_3 departing from 0 points to an asymmetry in the upwards and downwards fluctuations. Nonhealthy subjects are found to have larger values of $|c_2|$ and a positive c_3 (cf. Table IV), thus their temporal dynamics are characterized by stronger burstiness and a preponderance of upward fluctuations.

V. CONCLUSIONS

Sparse-SVM was used here to jointly optimize classification of healthy and nonhealthy FHR and select features amongst multifractal parameters that most significantly contribute to performance. Results clearly indicate that: i) multivariate classification outperforms univariate classification; ii) nonlinear multifractal features improve performance upon classification based on linear feature only ; iii) the use of a small values of p yields the best performance.

REFERENCES

[1] E. Chandraharan and S. Arulkumaran, "Prevention of birth asphyxia: responding appropriately to cardiotocograph (CTG) traces.," *Best Pract Res Clin Obstet Gynaecol*, vol. 21, pp. 609–624, Aug 2007.

[2] FIGO, "Guidelines for the use of fetal monitoring," *International Journal of Gynecology & Obstetrics*, vol. 25, pp. 159–167, 1986.

[3] M. Doret, H. Helgason, P. Abry, P. Gonçalvès, C. Gharib, and P. Gaucherand, "Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor," *Am. J. Perinatol.*, vol. 28, no. 4, p. 259, 2011.

[4] R. Leonarduzzi, J. Spilka, H. Wendt, S. Jaffard, M. E. Torres, P. Abry, and M. Doret, "p-leader based classification of first stage intrapartum fetal hrv," in *VI Congreso Latinoamericano de Ingeniería Biomédica (CLAIB)*, pp. 669–672, 2014.

[5] H. Wendt, P. Abry, and S. Jaffard, "Bootstrap for empirical multifractal analysis," *IEEE Signal Proc. Mag.*, vol. 24, no. 4, pp. 38–48, 2007.

[6] R. Leonarduzzi, H. Wendt, S. Jaffard, S. Roux, M. Torres, and P. Abry, "Extending multifractal analysis to negative regularity: p-exponents and p-leaders," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 305–309, 2014.

[7] H. Wendt, K. Kiyono, P. Abry, J. Hayano, E. Watanabe, and Y. Yamamoto, "Multiscale wavelet p-leader based heart rate variability analysis for survival probability assessment in chf patients," in *Proc. 36th Ann. Int. IEEE EMBS Conf.*, pp. 2809–2812, 2014.

[8] L. Laporte, R. Flamary, S. Canu, S. Déjean, and J. Mothe, "Nonconvex regularizations for feature selection in ranking with sparse svm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 6, pp. 1118–1130, 2014.

[9] M. Doret, M. Massoud, A. Constans, and P. Gaucherand, "Use of peripartum ST analysis of fetal electrocardiogram without blood sampling: a large prospective cohort study," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 156, no. 1, pp. 35–40, 2011.

[10] H. Gonçalves, A. P. Rocha, D. Ayres-de Campos, and J. Bernardes, "Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery," *Medical and Biological Engineering and Computing*, vol. 44, no. 10, pp. 847–855, 2006.

[11] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 1998.

[12] S. Jaffard, P. Abry, and H. Wendt, *Irregularities and Scaling in Signal and Image Processing: Multifractal Analysis*. World scientific publishing, 2015. to appear.

[13] B. Castaing, Y. Gagne, and M. Marchand, "Log-similarity for turbulent flows," *Physica D*, vol. 68, no. 3-4, pp. 387–400, 1993.

[14] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[15] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.