Scattering Transform for Intrapartum Fetal Heart Rate Variability Fractal Analysis: a Case-Control Study

Václav Chudáček[(1,4)], Joakim Andén[(2)], Stéphane Mallat[(2)] *Fellow, IEEE*, Patrice Abry[(1)] *Fellow, IEEE*, Muriel Doret[(3)]

Abstract—Intrapartum fetal heart rate monitoring, aiming at early acidosis detection, constitutes an important public health stake. Scattering Transform is proposed here as a new tool to analyze intrapartum fetal heart rate variability. It consists of a non linear extension of the underlying Wavelet Transform, that thus preserves its multiscale nature. Applied to a Fetal Heart Rate (FHR) signal database constructed in a French academic hospital, the Scattering Transform is shown to permit to efficiently measure scaling exponents characterizing the fractal properties of intrapartum fetal heart rate temporal dynamics, that relate not only to the sole covariance (correlation scaling exponent) but also to the full dependence structure of data (intermittency scaling exponent). Such exponents are found to satisfactorily discriminate temporal dynamics of Healthy subjects (from that of Non Healthy ones) and to emphasize the role of the *highest frequencies* (around and above $1\mathrm{Hz}$) in intrapartum fetal heart rate variability. This permits to achieve satisfactory classification performance, that improves on those obtained from the analysis of International Federation of Gynecology and Obstetrics (FIGO) criteria, notably by classifying as Healthy a number of subjects that were incorrectly classified as Non Healthy by classical clinically-used FIGO criteria. Combined to obstetrician annotations, these scaling exponents enables us to sketch a typology of these FIGO-False Positive subjects. Also, they permit to monitor the evolution along time of the intrapartum health status of the fetuses and to estimate an optimal detection time-frame.

Index Terms—Scattering transform ; Fractal ; Non linear ; Multiscale ; Intrapartum fetal heart rate variability ; Health status time evolution ; Classification.

I. INTRODUCTION

A. Motivation: Intrapartum fetal heart rate variability surveillance and acidosis detection

In delivery wards throughout the world, cardiotocograms (CTG) – combination of fetal heart rate (FHR) and uterine contraction signals [1] – are monitored with the aim to detect fetus hypoxia. Early detection enables obstetricians to act accordingly and reduce the subsequent fetal and neonatal mortality (cf. e.g., [1], [2]). In clinical routine, intrapartum

[(1)]   CNRS, ENS Lyon, Physics Dept., Lyon, France firstname.lastname@ens-lyon.fr
[(2)]   CNRS, ENS Paris, Math Dept., Paris, France firstname.lastname@ens.fr
[(3)]   Femme-Mère-Enfant Hospital, Lyon, France muriel.doret@chu-lyon.fr
[(4)] Czech Technical University in Prague, Czech Republic

surveillance essentially relies on the visual inspection of FHR signal, with a significant part of the final assessment relying on the evaluation of fetal heart rate variability (F-HRV). The health status of the fetus can be assessed using International Federation of Gynecology and Obstetrics (FIGO) guidelines [3], that essentially consist of a set of rules evaluating manually extracted temporal characteristics of CTG, such as baseline level, variability level, number and type of decelerations and their relation to contraction occurrence times (cf. [4]). Departures from normality, as defined by clinical guidelines, are regarded as sign of degradation of fetal normoxy and practically result in actions from obstetricians either aiming at improving the fetal state or at operating the delivery. While application of the FIGO rules permits high sensitivity to intrapartum acidosis detection, it also results in a poor specificity: Strict application of FIGO rules leads to a significant number of unnecessary operative delivery decisions, where post-birth exams a posteriori indicate that the fetus was efficiently coping with stress [4]. Operative deliveries may result in either immediate or long-term severe consequences for both the newborn and the mother. It thus constitutes an important public health stake to reduce the number of unnecessary operative deliveries, which motivates significant research efforts i) to automatize the computation of robust features from digitized CTG and ii) to produce sound statistical characterization of CTG and F-HRV beyond the essentially morphological (or geometric) FIGO criteria (cf. e.g., [5], [6] or [7], [8] for tentative reviews).

B. Motivation: Related works: Intrapartum fetal heart rate variability statistical characterization

To go beyond the clinically used static, temporal and pattern-based description of F-HRV, it has naturally been envisaged to use frequency based analysis. Following the seminal work in [9] for adult heart rate characterization, spectrum estimation tools were massively used for intrapartum F-HRV analysis (e.g., [10] and references therein). However, spectrum estimation for intrapartum F-HRV analysis suffers from important shortcomings. First, because of the time-evolving nature of the delivery process (baseline variations, occurrence of decelerations), intrapartum F-HRV signals are much less stationary than adult HRV signals are. This lead to the use of joint time-frequency representations to account for the time varying nature of the frequency content of intrapartum

F-HRV [11], or, along another line, to adaptive (or data-driven) characterization (cf. e.g., [12]). Second, spectrum estimation only captures the (second order statistics) correlation of data – thus neglecting higher statistical order forms of dependencies. Tremendous research efforts were devoted to overcome this limitation, often referred to as *non linear* analysis in the literature, to explicitly emphasize that the capture of information beyond (linear) correlation is intended [13], [14], [6]. Entropy rates, stemming from dynamical system modeling, and probing partially the joint distribution function, thus higher order statistics, have also been used (cf. e.g., [15]). Alternatively, to describe F-HRV beyond correlation, the point process nature of heart beat has been fruitfully explored [16]. The fractal, or scaling, paradigm has also been involved into F-HRV analysis [14], [17] and recently, multifractal analysis has been shown to permit an efficient measure of statistical dependence beyond the sole correlation [18], [19], [20].

C. Motivation: Goals and Contributions

The present contribution intends to explore the benefits of using the Scattering Transform for intrapartum F-HRV analysis. It consists of a recently proposed non linear and multiscale transform [21], shown to be highly effective to classify audio signals, image textures, and to analyze fractal properties [22], [23]. The Scattering Transform is defined in Section II. It is applied to a FHR database, described in Section III, carefully constructed and well documented by obstetricians at the *Femme-Mère-Enfant* (Woman-Mother-Child) academic public Hospital (HFME), in Lyon, France. It is shown, first, how the Scattering Transform enables to capture and quantify the fractal properties of intrapartum F-HRV data, and how the extracted (correlation and intermittency) scaling exponents enable to distinguish subjects suffering from acidosis from healthy ones, and how these scaling exponents emphasize the role of the *highest frequencies* (around and above 1Hz) in F-HRV temporal dynamics (cf. Section IV-A). Classification performance are then quantified and compared against FIGO-based achieved ones (cf. Section IV-B). Further, it is shown how these scaling exponents enable to track the evolution along time of the fetus health status from Healthy to Non Healthy (cf. Section IV-C). Finally, making use of the documentation provided by the obstetricians, a typology of the subjects misclassified as Non Healthy using the FIGO-rules (FIGO-False Positive), compared to the classification achieved using Scattering Transform based scaling exponents, is performed in Section IV-D. This contribution elaborates on a preliminary work presented at EMBC2013 [24].

II. METHODS

A. Scattering Transform

A scattering transform provides locally translation invariant multiscale coefficients, which characterize the scaling properties of signals. They are computed by iteratively calculating the modulus of complex wavelet coefficients [21], [22], [25]. Let $X(t)$ denote the time series to analyze. The wavelet $\psi(t)$ is a complex analytic band-pass filter, whose transfer function is thus supported over positive frequencies. Let $\psi_j(t) =$

$2^{-j}\psi(2^{-j}t)$ denote the dilated templates of ψ at scales $a = 2^j$. While the wavelet transform computes $X \star \psi_j(t)$ for multiple scales 2^j, the scattering transform outputs locally translation invariant coefficients by averaging the modulus of these complex coefficients. Let $\phi(t)$ be a low-pass filter, which is dilated to adjust the averaging support: $\phi_J(t) = 2^{-J}\phi(2^{-J}t)$. The first order scattering coefficients are thus defined as the average amplitude of wavelet coefficients, for any $1 \leq j \leq J$, over half overlapping time windows of size 2^J, centered at the points $t = k2^{J-1}$, $k \in \mathbb{N}$:

$$SX(j,k) = |X \star \psi_j| \star \phi_J(t = k2^{J-1}) \ . \qquad (1)$$

Averaging results in the loss of the high frequency contents of $|X \star \psi_{j_1}(t)|$, which can be recovered by computing a new set of wavelet coefficients $|X \star \psi_{j_1}(t)| \star \psi_{j_2}(t)$. Averaging their modulus defines the second order scattering coefficients at each $t = k2^{J-1}$, for any $1 \leq j_1 < j_2 \leq J$:

$$SX(j_1, j_2, k) = ||X \star \psi_{j_1}| \star \psi_{j_2}| \star \phi_J(t = k2^{J-1}) \ . \qquad (2)$$

This iterative procedure can be extended to higher orders. Third order coefficients are similarly defined for any $1 \leq j_1 < j_2 < j_3 \leq J$ by $SX(j_1, j_2, j_3, t) = |||X \star \psi_{j_1}| \star \psi_{j_2}| \star \psi_{j_3}| \star \phi_J(t)$. Only the two first order coefficients are exploited in the present work.

By definition, the amplitude of second order coefficients depends upon that of the first order coefficients. To remove such dependency, it is convenient to introduce the normalized second order scattering coefficients:

$$\widetilde{S}X(j_1, j_2, k) = \frac{SX(j_1, j_2, k)}{SX(j_1, k)} \ . \qquad (3)$$

In the sequel, the vector of scattering coefficients aggregates, for each time position k, the first and normalized second order coefficients:

$$\mathcal{S}X(k) = \Big(\{SX(j,k)\}_{1 \leq j \leq J}, \{\widetilde{S}X(j_1, j_2, k)\}_{1 \leq j_1 < j_2 \leq J} \Big) \ . \qquad (4)$$

While providing a multiscale representation of X, the scattering transform consists of a highly non linear transform, as opposed to the underlying discrete wavelet transform.

In practice, in the present contibution, a complex wavelet is used, consisting of the analytic part (restriction to positive frequencies) of a Battle-Lemarié cubic spline wavelet [21]. The window ϕ is the cubic spline scaling function associated to this wavelet. The ScatNet software is available at http://www.di.ens.fr/data/software/scatnet/.

B. Fractal Dynamics

When computed from time series X possessing fractal dynamics, or scale invariant, properties, as well as stationary increments, the scattering coefficients have been shown [25] to exhibit power law behavior with respect to scales:

$$SX(j,k) \quad \simeq \quad 2^{jz_1(k)} \ , \qquad (5)$$
$$\widetilde{S}X(j_1, j_2, k) \quad \simeq \quad 2^{(j_2-j_1)z_2(j_1,k)} . \qquad (6)$$

As first order coefficients average amplitude wavelet coefficients, z_1 directly provides an estimate of the Hurst parameter

H, for fractional Brownian motion and is in general related to the covariance function of data, it is thus referred to as the *correlation scaling exponent*. The normalized second order scattering coefficients provide an information beyond 2nd order statistics [25]. Furthermore, for exactly scale invariant processes, $z_2(j_1)$ has been shown to be independent of j_1: $z_2(j_1) \equiv z_2, \forall j_1$ [25]. For fractional Brownian motion, for instance, $z_2 = -1/2$. For multifractal processes, z_2 has been shown to be related to the *intermittency* or *multifractality* parameter. In general, the $z_2(j_1)$ captures the bursty or intermittent nature of data: The larger the $z_2(j_1)$ the more bursty in time the data. The $z_2(j_1)$ are thus referred to as the *intermittency scaling exponents*.

The scaling exponents z_1 and $z_2(j_1)$ thus provide features characterizing the fractal dynamics of X, where z_1 gives a global regularity information which mostly depends upon its second-order statistics, whereas the $z_2(j_1)$ depend upon higher-order statistics. Fractal properties in time series, in general, and in HRV in particular, can also be analyzed using other tools such as e.g., those developed and used in [18], [19] . However, while the Scattering Transform is not primarily intended for fractal analysis, but rather for non linear analysis, it also formally offers an original manner to measure fractal properties in data via non linear transformations. Such connections are made explicit in [25]. While theoretically formally equivalent when applied to synthetic truly fractal processes, the scattering based measurements of fractal property on real world data may potentially differ from those obtained from other fractal estimation tools.

C. Data preprocessing and Scattering Transform computation

As common practice for HRV analysis (cf. e.g., [14], [26]), the lists $\{t_n\}_{n=1,...,N}$ of R-peaks are transformed into regularly sampled beat-per-minute (BpM) time series, $X(t)$, by linear interpolation of the measurements $\{(t_n/1000, 60000/(t_{n+1} - t_n))\}_{n=1,...,N}$. As F-HRV carries by nature no information beyond 3 Hz, sampling frequency is set to $f_s = 8$ Hz, (using higher f_s has been observed to yield no improvement in classification).

To be able to follow the evolution along time of the health status of the fetuses, the scattering transform is computed in T-minute long sliding windows. For clinical practice, obstetricians expect regular and short updates on the fetus health status, with a typical update period of 5 to 10 min. Thus, for this study, $J = 12$ is used so that $2^J/f_s = 512$s $\simeq 8.5$min (as the optimal use of the current version of the scattering transform used here requires power of 2 sample size), with 50% overlap. Even though the database consists of only 45 subjects, this sliding time-window analysis procedure amounts to computing scattering coefficients, $\mathcal{S}X(k)$, for 507 different time windows, indexed by k, for each subject.

While studying the evolution along time, k, of the statistics of $\mathcal{S}X(k)$ enables us to follow the time evolution of the fetus health status, an average performed on the **last-K-windows** before delivery can be assumed to measure the fetus health status before delivery, and hence at the time when obstetricians make the decision to operate delivery or not.

By construction of the BpM time series (interpolation at 8 Hz), octave j_1, corresponding to frequencies ranging from 2 to 4Hz, contains no or little information related to F-HRV temporal dynamics and is thus discarded from analysis.

III. DATABASE

A. Data measurement

Intrapartum CTG have routinely been monitored at HFME over the last 30 years, with systematic STAN-based surveillance for fetuses suffering from initial intermediate FHR during labor or with high risk of fetal asphyxia (post-date delivery, intra-uterine growth restriction, diabetes, occurrence of abnomalies in CTG, . . .). CTGs are measured using the STAN, Neoventa Medical (Moelndal, Sweden) system (STAN 21 or 31 systems, 12bit resolution, 500Hz sampling rate for the FECG signal), thus producing high quality data compared to the less invasive but far less reliable ultrasound doppler-effect based measurements. From CTG measurements, for each subject, a list, $\{t_n\}_{n=1,...,N}$, of beat-by-beat R-peak occurrence time (in ms) is available.

B. Database

Obstetricians have carefully selected subjects and annotated files according to FIGO guidelines to create a documented database. The following criteria were used for inclusion of a subject into the database: The database must contain *representative* Healthy and Non Healthy subjects, which were correctly diagnosed by FIGO-rules as such, as well as representative Healthy subjects which were incorrectly diagnosed as Non Healthy by FIGO-rules; For each subject, the F-HRV time series are at least 30 minute long, many recordings last for several hours; Data have *good quality*, i.e., there are in general few outliers or missing beats in the R-peak occurrence time list; Recordings are *well-documented* by obstetricians. The documentation include umbilical cord artery acid-base (pH) status describing fetal acid-status at delivery, AGPAR score, delivery mode, delay from end-of-recording to delivery time. It also includes description of further obstetrician motivations for operative delivery, related to the occurrence of final brady-cardia, shape of decelerations and their delay with respect to contractions, level of variability and reactivity.

The database is organized into 3 classes:
i) **FIGO-TN:** 15 fetuses with normal fetal outcome (defined as Apgar score of 10 at 5 minutes of life and arterial umbilical cord pH > 7.30, hence non acidotic thus healthy) and CTG classified as normal, thus referred to as FIGO-True Negatives ;
ii) **FIGO-TP:** 15 fetuses with fetal acidosis (arterial umbilical cord pH < 7.05, hence abnormal) and CTG classified as abnormal (hence correctly diagnosing fetal state as abnormal), thus referred to as FIGO-True Positives ;
iii) **FIGO-FP:** 15 fetuses with normal fetal outcome (Apgar score of 10, arterial umbilical cord pH > 7.30, hence healthy), yet with pathological CTG (hence incorrectly diagnosed as abnormal), thus referred to as FIGO-False Positives.

These 3 classes correspond to only two groups for fetal health status: **Non Healthy**, equivalent to the FIGO-TP class; and **Healthy** corresponding to the union of the FIGO-TN and

FIGO-FP classes.

This three-class database provides us with a FIGO reference benchmark reported in Table II. The goal of this case study analysis is twofold: To improve performance, by decreasing the number of false positives and ; To produce a typology of these false positives to analyze why they were misclassified.

IV. RESULTS

A. Fractal dynamics and scattering transform: scaling range and scaling exponents as discriminating features
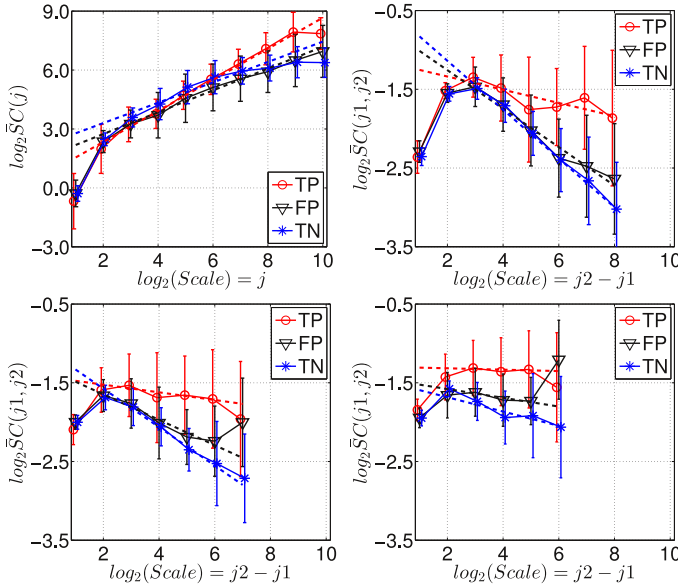

Fig. 1.   **Fractal Dynamics.** $\log_2 \overline{S}C(j)$ versus $\log_2 2^j = j$ (top left), and $\log_2 \overline{S}C(j_1, j_2)$ versus $\log_2 2^{j_2-j_1} = j_2 - j_1$ for $j_1 = 2$ (top right), $j_1 = 3$ (bottom left) and $j_1 = 4$ (bottom right).

1) Fractal dynamics and scattering transform: It has often been argued in the literature that fractal temporal dynamics constitutes a relevant and fruitful model for F-HRV modeling (cf. e.g., [14], [19], [17]). Eqs. 5 and 6 in Section II-B indicate that, for time series with fractal dynamics and stationary increments, the scattering coefficients are expected to show power-law behaviors with respect to scales. To investigate such behaviors on F-HRV BpM time series, let us compute $\overline{S}C(j)$ and $\overline{S}C(j_1, j_2)$, corresponding respectively to the averages, per class, for each class $C = TP, TN, FP$, of the first order $SX(j, t)$ and normalized second order $\overline{S}X(j_1, j_2, t)$ scattering coefficients, over the last-3-windows. The choice $K = 3$, which (given the overlap) corresponds to the last 17min before delivery, and hence matches the decision time frame in clinical situation, is further justified in Section IV-C.

Fig. 1 displays $\log_2 \overline{S}C(j)$ as a function of $\log_2 2^j = j$ and $\log_2 \overline{S}C(j_1, j_2)$ as a function of $j_2 - j_1$, for different j_1 (95% confidence intervals, computed from within-class standard deviations are superimposed). Linear behaviors in such log-log plots (superimposed dashed lines) indicate that power law behaviors such as those modeled in Eqs. 5 and 6 are observed in F-HRV BpM time series, and hold on average across octaves $3 \leq j \leq 9$ for $\overline{S}C(j)$ (top left plot) and for $3 \leq j_2 - j_1 \leq J - 2 - j_1$ for $\overline{S}C(j_1, j_2)$. These observations

validate the relevance of the concept of fractal to describe F-HRV temporal dynamics across time scales ranging from 1s $\leq 2^j \leq$ 1min approximately (or equivalently for frequencies in $0.01 \leq f \leq$ 1Hz). That range encompasses and slightly enlarges the frequency range involved into the Low-Frequency vs. High-Frequency band decomposition, classically used for adult HRV analysis [9], [14] and much debated in the context of intrapartum F-HRV (cf. e.g.,

Fig. 1 also clearly evidences that the $\log_2 \overline{S}C(j_1, j_2)$, as functions of $j_2 - j_1$ do not overlap when computed for various j_1, and thus that $z_2(j_1)$ do depend on j_1, in contradistinction with the theoretical results in [25] that show that for processes with exact scale invariance and stationary increments $\overline{S}C(j_1, j_2)$ and $z_2(j_1)$ should not depend on j_1. These empirical observations clearly indicate that the $z_2(j_1)$s for different j_1s do not probe the same information beyond correlation and also fractal constitutes only a global and approximate model for F-HRV temporal dynamics, rather than a strictly exact one.

These empirical observations (first, scaling behaviors of the scattering coefficients over scales that range from $0.01 \leq f \leq$ 1Hz ; second, departures from exact scale invariance) suggest to measure systematically the scaling exponents $\{z_1(k), z_2(j_1, k), j_1 = 2, 3, 4\}$, for each subject of each class and for each time window k, and for different j_1, and to investigate their potential as discriminating features for acidosis detection.

2) Scaling exponent estimation: Estimation of the scaling exponents $\{z_1(k), z_2(j_1, k), j_1 = 2, 3, 4\}$, per subject and per time window, is achieved by linear regressions in $\log_2 SX(j, k)$ vs. j diagrams (for $3 \leq j \leq 8$) and in $\log_2 \tilde{S}X(j_1, j_2, k)$ vs. $j_2 - j_1$ (for $3 \leq j_2 - j_1 \leq J - 3 - j_1$.

TABLE I
Discrimination. P-VALUES FROM WILCOXON RANKSUM TEST FOR THE NULL HYPOTHESIS OF EQUALITY IN MEAN OF THE DISTRIBUTIONS.

	TP vs. TN	TP vs. FP	FP vs. TN	nonH vs. H
z_1	**<0.001**	**<0.001**	0.320	**<0.001**
$z_2(j_1 = 2)$	**<0.001**	**<0.001**	0.263	**<0.001**
$z_2(j_1 = 3)$	**0.038**	0.967	**0.009**	0.243
$z_2(j_1 = 4)$	0.901	**0.047**	**0.010**	0.284

3) Discriminating power of z_1: Fig. 2 compares, by means of BoxPlots, for the last-3-windows, the distributions per class (TP, FP, TN) or health status (Non healthy; Healthy) of the correlation scaling exponents z_1 and $z_2(j_1)$. It is complemented by Table I, reporting the p-values obtained using the *Wilcoxon ranksum test*, with equality in mean of the distributions as null hypothesis. Both Fig. 2 and Table I clearly indicate that scaling exponent z_1 efficiently discriminates Healthy from Non Healthy, and further also distinguishes the three pairs of classes, with the distributions of z_1 for the TN and FP classes being much closer than they are from that of the TP class. This is a clear indication that z_1 sees the **FIGO-FP** as closer to the **FIGO-TN**, than to the **FIGO-FP** and thus actually as Healthy subjects. This is in close agreement with what has been observed using Hurst parameter, or some multifractal attributes as discriminative features (cf. [27], [19], [17]).
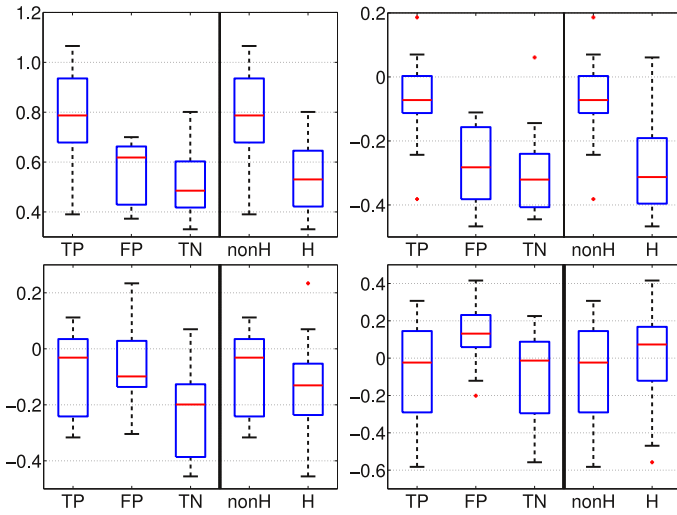
Fig. 2. **BoxPlots** comparing, for the last-K-windows, the distributions per class (TP, FP, TN) and group (Non H, H) of the scaling exponents z_1 (top left) and $z_2(j_1)$ for $j_1 = 2$ (top right), $j_1 = 3$ (bottom left) and $j_1 = 4$ (bottom right). Central marks and box edges correspond to median and 25th-75th percentiles, whiskers indicate extreme values that are not considered outliers, while outliers are plotted individually.

4) Discriminating power of $z_2(j_1)$: Fig. 2 and Table I show that the intermittency scaling exponent $z_2(j_1 = 2)$ achieves a good discrimination between Healthy and Non Healthy and between the three pairs of classes, and that the distributions of $z_2(j_1 = 2)$ for the TN and FP classes are much closer between each other than they are from the distribution for TP class. Interestingly, for $z_2(j_1 = 3)$ and $z_2(j_1 = 4)$, conclusions differ as those two scaling exponents do not discriminate neither Healthy from Non Healthy subjects nor the FP from the TP classes, which constitutes the contribution major target.

This clearly indicates that the $z_2(j_1)$ for different j_1 do not probe the same information, and thus confirms that F-HRV BpM fractal dynamics are only approximately (and not exactly) described by exact scale invariant processes.

5) Frequency band analysis: This analysis can be made more precise in terms of frequency bands. Octave $j_1 = 2$ represents frequencies ranging from 1 to 2Hz, i.e., around and above 1 Hz (or equivalently time scales ranging from 0.5 to 1s), which correspond to the *highest frequencies* contributing to F-HRV temporal dynamics ; $z_2(j_1 = 2)$ thus measures the temporal dynamics beyond correlation specifically attached to that range that will be referred to as the *highest frequency*, that thus turns out to be crucial to discriminate Healthy from Non Healthy temporal dynamics. Note that octave $j_1 = 2$ corresponds to frequencies beyond the traditional *High Frequency* band, stemming from the adult-devised *High Frequency/Low Frequency* band splitting, corresponding respectively to $[0.15, 0.40]$ and $[0.04, 0.15]$ Hz. Conversely, Octaves $j_1 = 3$ and $j_1 = 4$, ranging respectively from 0.5 to 1Hz and 0.25 to 0.5Hz, essentially match the *High Frequency* band. The dependence structure for those frequency ranges, measured by $z_2(j_1 = 3)$ and $z_2(j_1 = 4)$, is found to be non discriminative between Healthy and Non Healthy.

Interestingly, this indicates that dependence information

beyond correlation, relevant to discriminate between the temporal dynamics of Healthy and Non Healthy fetuses, must be associated to the *highest frequencies* around and above 1 Hz of F-HRV, beyond the classical HF band. This provides us with new lights on the temporal dynamics of F-HRV.

From now on, further analyses are therefore focused on the two scaling exponents, z_1 and $z_2(j_1 = 2)$, as they show the largest powers in discriminating Healthy from Non Healthy and FIGO-TP from FIGO-FP.

6) Interpretations in terms of frequency contents of F-HRV and Good Variability*:* The boxplots in Fig 2, as well as the scatterplot of $z_1, z_2(j_1 = 2)$ for the last-3-windows, in Fig. 3 (middle plot), reveal that z_1 and $z_2(j_1 = 2)$ for Healthy subjects takes systematically lower values compared to those of Non Healthy subjects.

From Eq. 5, lower z_1 indicate larger contributions of high frequencies (or fine scales) to F-HRV temporal dynamics, compared to low frequencies (coarse scales) (where fine scales refer to $j_1 = 3$, hence to $a \simeq 1s$ and to $f \simeq 1$Hz, and coarse scales to $j_1 = 8$, hence to $a \simeq 1$min or $f \simeq 0.015$Hz). This High Frequency dominant contribution for Healthy subjects is in agreement with earlier results obtained using multifractal analysis (cf. [19], [17]).

Exponent $z_2(j_1 = 2)$, computed from normalized second order scattering coefficients and from Eq. 6 requires more subtle interpretations: First, Figs 1 and 2 unambiguously show that the $\log_2 \overline{S}C(j_1, j_2)$, and thus the $z_2(j_1 = 2)$ for $j_1 = 2$, computed from the Non Healthy class, are systematically larger than those obtained from the Healthy FIGO-FP subjects, themselves larger that those produced by the Healthy FIGO-TN subjects. This indicates that the temporal dynamics of Non Healthy fetuses are more intermittent and bursty that of Healthy fetuses. This can naturally be expected generically via the interpretation that a biological system under stress likely yields complicated reactions, materialized by bursty and intermittent temporal dynamics. Second, resulting from the dependence structure beyond the simple correlation structure, $z_2(j_1 = 2)$ measures the way the energy at high frequencies (around 1 and 2Hz) is modulated along time: Smaller $z_2(j_1 = 2)$ for Healthy fetuses thus also indicate that such modulations occur at higher frequency than they do for Non Healthy subjects.

Such analyses renew the notion of *good variability*: Instead of being defined as the amplitude of the fluctuations of the BpM time series, at a a priori chosen scale, being larger than a given threshold, it is proposed here that *good variability* can be measured via the fractal paradigm, i.e., by the way such fluctuations vary from one scale to the other, or, in other words, by a scaling exponent. In this framework, good variability is assessed by low correlation and intermittency scaling exponents, that indicate a larger contribution of high frequencies to temporal dynamics. High frequencies and low frequencies are empirically (and not a priori) defined as the upper and lower limits of the observed scaling behaviors matching Eqs. 5 and 6 and corresponding to time scales of the order of respectively, one second (or 1Hz) and one minute (or 0.015Hz).
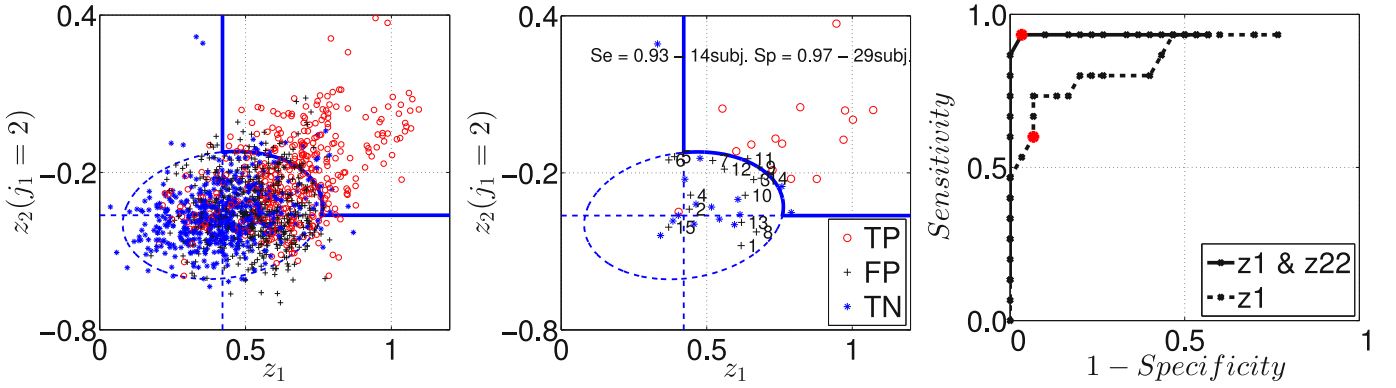
Fig. 3. **Classification.** Left, scatterplot of $z_1, z_2(j_1 = 2)$ for the all time windows, for the three classes (FIGO-TP: red 'o', FIGO-FP, black '+', FIGO-TN, blue $*$), the spontaneous fluctuations of the FIGO-TN scaling exponents defines the *Healthy Domain* as the union of an ellipse level line, corresponding to the largest achieved MCC, of the best fit bivariate Gaussian and $z_1 \leq T_1$, $z_2(j_1 = 2) \leq T_2$ zones ; Middle, scatterplot of the median of $z_1, z_2(j_1 = 2)$ for the last-K-windows (black numbers correspond to the identifiers s of the FIGO-FP subjects, as used in the typology in Section IV-D); ROC curve obtained by varying ellipse level line, with best MCC shown as a (red) 'o' (and compared to the ROC stemming from the use of the sole z_1).

B. Classification performance

From the scatterplots in Fig. 3, it can be observed i) that z_1 and $z_2(j_1 = 2)$ are, for the last-3-windows (middle plot) systematically larger for Non Healthy subjects that thus live in the upper right corner of the $z_1, z_2(j_1 = 2)$ plan ; and ii) that the joint distribution of $z_1, z_2(j_1 = 2)$ for the FIGO-TN class, for all time windows (left plot), can be well modeled by a bivariate Gaussian law. This lead us to define the *Healthy Domain* as the inside of the ellipse corresponding to the level line of the fitted bivariate Gaussian law that maximizes the Matthews correlation coefficient (MCC) of a classification performed using the last-3-windows, complemented with the union of the portions of the plan delimited by $z_1 \leq T_1$, $z_2(j_1 = 2) \leq T_2$, with T_i defined as the mean of z_i for the FIGO-TN class. That definition of the *Healthy Domain* as well as the choice $K = 3$ is further justified in Section IV-C2.

From that definition of the *Healthy Domain*, each time window can be classified as Healthy or Non Healthy. Then a per subject majority vote procedure classifies each subject as Healthy or Non Healthy. This is illustrated in Fig. 3 (center plot) which shows the median last-3-window position for each of the 45 subjects, compared to the Healthy Domain. A Receiver Operational Characteristic (ROC) curve can be computed by varying the bivariate Gaussian distribution elliptic level line. It is plotted in Fig. 3 (left plot) and shows first that the ROC curve obtained from using jointly z_1 and $z_2(j_1 = 2)$ exhibits systematically better performance than that obtained from the use of the sole z_1 (see also Table II). This clearly validates the empirical observation reported along Section IV-A: $z_2(j_1)$, related to higher order dependence structure, captures temporal dynamics features that are not already captured in z_1 (related to the sole 2nd-order correlation) and illustrates the benefits on the non-linear nature of the scattering transform. Fig. 3 (left plot) and Table II show, second, that the scattering transform based $z_1, z_2(j_1 = 2)$ exponent classification procedure benefits from excellent performance, and, at optimum (i.e., for the largest MCC), yields significant improvements compared to results obtained from a FIGO-criteria based classification.

TABLE II
Performance: SPECIFICITY, SENSITIVITY, POSITIVE PREDICTIVE VALUE, F-MEASURE [28]) AND MATTHEWS CORRELATION COEFFICIENT [29].

	Se	Sp	PPV	F	MCC
$FIGO$	1.00	0.50	0.50	0.67	0.50
z_1	0.60	0.93	0.82	0.69	0.59
z_1 & $z_2(j_1 = 2)$	**0.93**	**0.97**	**0.93**	**0.93**	**0.90**

C. Time evolution

1) Sample paths in the $z_1, z_2(j_1 = 2)$ plan: So far, focus has been only on the last-3-windows before delivery. However, the scaling exponents $z_1(c, s, k), z_2(c, s, k, j_1 = 2)$ can be studied as functions of time k, to analyze the time evolution of the fetus health status. A Healthy subject remains Healthy from the beginning to the end of the recording, thus the corresponding sample path remains quasi exclusively within the Healthy Domain, as illustrated in Fig. 4, left plot. Conversely, for long enough recordings, a Non Healthy subject starts evolution in the Healthy domain and moves outside after a certain time, as shown in Fig. 4, middle left plot. This can also be seen on the scatterplot gathering the 507 time windows for the 3 classes: A large number of the Non Healthy subject early time windows (when the subject is presumably still healthy) sit in the Healthy Domain, while the late windows are outside. For the FIGO-FP that are correctly identified as Healthy by the scaling exponents $z_1, z_2(j_1 = 2)$, the corresponding sample paths quasi-continuously remain within the healthy domain, as illustrated in Fig. 4, middle right plot. For some FIGO-FP that are either not correctly identified as Healthy by the scaling exponents $z_1, z_2(j_1 = 2)$ or close to the border, the corresponding sample path often leaves temporarily the Healthy domain, before returning into it and leaving it again, as illustrated in Fig. 4, right plot. Sample paths for all subjects are available at perso.ens-lyon.fr/patrice.abry/SamplePaths.pdf.

These time evolutions can be further quantified. On average, the FIGO-TP spend more than 50% of their time outside the Healthy Domain, against less than 5% for the FIGO-TN and FIGO-FP. Moreover, the FIGO-TP spend, on average, 6 time
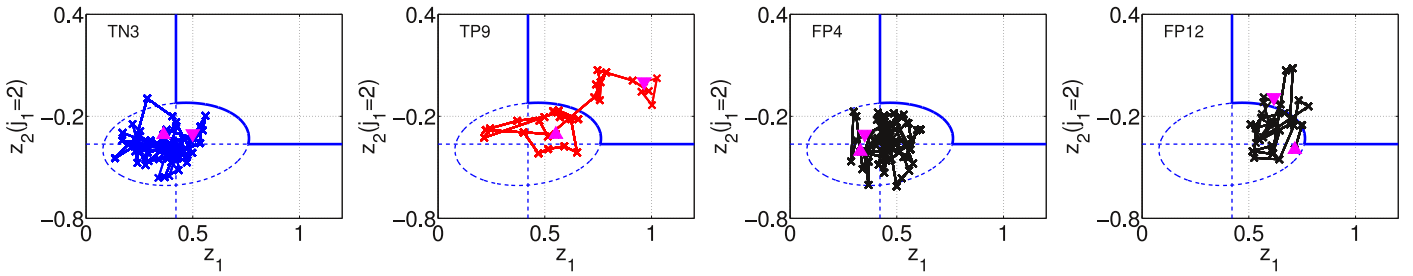
Fig. 4. **Sample paths in** $z_1, z_2(j_1 = 2)$ **plan.** A Healthy subject (top left) performs a random walk in the Healthy Domain ; A Non Healthy subject starts in the Healthy Domain but at some time leaves it definitely towards the upper left corner (top right) ; A FP subject correctly identified as Healthy by the scaling exponents also remains in the Healthy Domain (bottom left) ; while a FP subject incorrectly classified as Non healthy by the scaling exponents oscillates around the border of the Healthy Domain. Upper and lower triangles mark respectively the beginning and end of the sample paths.

windows, out of the last 7 (last $\simeq$ 35 minutes before delivery) outside the Healthy domain, against less than 1 for the FIGO-TN and FIGO-FP. This clearly underlines a significant difference in the health status time evolution between the FIGO-TP and FIGO-FP classes: While FIGO-TP subjects remain for long periods of time outside of the Healthy Domain, FIGO-FP subjects leave it only occasionally and for short periods of time, before returning into it and possibly leave it again briefly. The sample path of the scaling exponents $z_1, z_2(j_1 = 2)$ in the corresponding 2D plan, can thus be considered as a time evolving diagnostic tool for obstetricians.

2) Last-K-Windows and Healthy Domain: ROC curves were computed from the last-K-windows vote procedure, for different choices of K, $1 \leq K \leq 41$, and compared in Fig. 5, left plot. This clearly shows that optimal classification performance are obtained for $K = 3$ (hence justifying that choice for the design of the Healthy Domain). This indicates that the last 17-minutes before delivery provide an optimal time frame for scaling exponent based decision making. The ellipse corresponding to the largest MCC for $K = 3$ has thus been selected to define the Healthy Domain used for classification. Fig. 5, right plot, shows how classification performance degrade for other choices of K. Variance in estimation of the scaling exponent is likely large enough to explain that the use of the sole last window, $K = 1$, performs worse than $K = 3$, despite corresponding to a shorter time before delivery. For $K = 5$ (last 25 minutes), performance degrade only slightly (compared to $K = 3$), in agreement with the fact that FIGO-TP are found consistently Non Healthy as early as 6 windows before delivery (cf. Section IV-C1). When K is further increased, Sensitivity and MCC are monotonously decreasing, a direct consequence of the fact that during early windows, Non Healthy subjects are actually still Healthy, thus the majority vote procedure involving early windows places Non Healthy subjects into the Healthy Domain.

This time evolution analysis of $z_1, z_2(j_1)$ sample path constitutes one of the rare quantitative measure of the departure from Healthy to Non Healthy fetus and of the corresponding departure time. Also, it quantifies objectively the optimal time frame, $K = 3$ to 5, corresponding to 17 to 25min, within which fractal dynamics must be measured to assess the health status of fetuses and to permit an efficient detection of Non Healthy subjects. An optimal time varying classification can

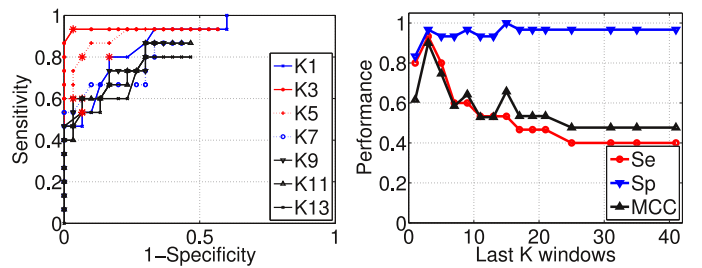thus be constructed using a $K = 3$ majority vote procedure.

Fig. 5. **Time Evolution.** Left, ROC curves for different Ks, red '∗' indicate for each K the largest MCC. It shows that $K = 3$ yields optimal performance. Right, performance (Sensitivity, Specificity and MCC) as functions of K.

D. Typology for False-Positive subjects

Beyond the classification performance figures reported in Table II, this database can be further analyzed using the annotations provided by obstetricians, so as to determine which FIGO-FP are actually correctly reclassified as Healthy by scaling exponents $z_1, z_2(j_1 = 2)$.

Obstetrician annotations indicate that **FIGO-FP** subjects FP1, FP2, FP4, FP7, FP8, FP10, FP12, FP13 and FP15 were classified pathological by FIGO-rules because of a *Long-period of Low Variability*, which precisely means that the corresponding BpM time series show a variability of less than 5 BpM over at least 20 minutes. The scatterplot of the scaling exponents $z_1, z_2(j_1 = 2)$ in Fig. 3, left plot, shows that all such subjects remain in the Healthy Domain and are thus classified correctly as Healthy. Obstetrician annotations also indicate that **FIGO-FP** subjects FP2, FP3, FP4, FP5, FP7, FP8, FP10, FP12 and FP15 were classified pathological because of *Low Reactivity*, that is a weak reactivity of the fetus heart after decelerations induced by contractions. Again, the scatterplot shows that these subjects are correctly classified as Healthy using the scaling exponents $z_1, z_2(j_1 = 2)$. This illustrates that subjects annotated by obstetricians as suffering from either low variability or low reactivity, essentially because they present BpM time series with low amplitudes in fluctuations, may actually very well exhibit temporal dynamics that very much resemble that of Healthy subjects, rather than that of Non Healthy ones. Interestingly, this confirms that scaling exponents taking low values constitute a measure of *good*

variability, that turns out to be more relevant and robust than the sole thresholding of the amplitude of the BpM time series fluctuations. Also, it is interesting to note that for low variability subjects (FP2, FP5, FP7, FP8, FP12, FP15), the sole exponent z_1, taking a low value, is enough to classify them as Healthy, while for low reactivity subjects (FP1, FP3, FP10, FP13), z_1 takes a large enough value to match the Non Healthy subject values. However, for these subjects, exponent $z_2(j_1 = 2)$ takes a low value thus maintaining them into the Healthy domain. There is hence a clear benefit in using jointly the scaling exponents z_1 and $z_2(j_1 = 2)$.

Some FP subjects remain incorrectly classified by the scaling exponents z_1 and $z_2(j_1 = 2)$ (FP9, FP11 and FP14). For those subjects, obstetrician annotations indicate heart rate decelerations, which are labeled either as *complicated-shape*, or as *deep* or as *delayed after contraction*. Also, FP3, FP10 and FP12 (close to Healthy Domain border) are indicated to suffer from *delayed after contraction* decelerations. Earlier studies reported in [19] or [12] on the same database show that these same subjects were also either not correctly classified or close to the border, using multifractal attributes or adaptive complexity measures as features.

These observations tend to suggest that the occurrence of *complicated-shape*, or *deep* or *late* decelerations in F-HRV accompanies an actual change in the temporal dynamics of the BpM time series, which is thus *feeled* by the scaling exponents. This change tends to occur jointly on z_1 and on $z_2(j_1 = 2)$ and thus affects the entire dependence structure of the BpM time series, and not only their sole correlation structure. Also, this change corresponds to a reduction of the contribution of high frequencies, in a manner that tends to resemble the temporal dynamics of Non Healthy subjects. Therefore, Healthy subjects presenting such types of decelerations have undergone a change in their temporal dynamics that corresponds, though less pronounced, to that of Non Healthy subjects. They are thus less easy to disentangle from Non Healthy subjects.

V. DISCUSSIONS, CONCLUSIONS AND PERSPECTIVES

The potential of the two correlation and intermittency scaling exponents measured from scattering transforms to characterize intrapartum fetal heart rate variability has been explored. Such exponents constitute quantitative measures of the fractal nature of BpM time series temporal dynamics, with the particular property, that they explore temporal dynamics, beyond the sole correlation level (or second statistical order), via the entire dependence structure of data (at all statistical order) levels.

These scattering transform based scaling exponents permit to first confirm that fractal dynamics is characteristic of F-HRV BpM signal, in frequencies ranging from 10 mHz to 1Hz, but also that exact scale invariance processes (such as fractional Brownian motions of multifractal) constitute only approximate models. Second, they enable to show clear differences between the temporal dynamics of Healthy and Non Healthy fetuses (for the former, high frequencies contribute more than for the later), and thus to renew the practical measure of *good variability*. Further, computed from sliding short time windows,

the scaling exponents z_1 and $z_2(j_1 = 2)$ enable to characterize the evolution along time of the fetus health status, to visualize when and how fetuses depart from the Healthy Domain, thus providing obstetricians with a potentially interesting tool to assist forming diagnostics. Combined to obstetrician annotations, these scaling exponents also enable us to draw a typology of FIGO-FP subjects.

The results obtained from this case study and documented database are promising and the present study will continue along different lines, under current investigations. At the methodological level, features extracted from scattering coefficients will first be compared, both in terms of nature (what they actually measure in data) and of performance (how well they classify), against other classical, or less classical, linear and mostly non linear features (FIGO-criteria; entropy-rate [15]; fractal and multifractal [27], [19], [17], ...). Notably, attention will be focused on whether, besides overall classification performance, the misclassified subjects are always the same or differ, when using different types of features. Secondly, it will be investigated whether the use of the entire collection of (54 per window) scattering coefficients into supervised machine-learning type classifiers (such as SVM) yields better or complementary classification performance, compared to those achieved from the sole two scaling exponents only, retained to achieve a non-supervised classification. At the practical level, the results obtained on this case-study database will be comforted and complemented on the large database currently been constituted at HFME (above 3000 subjects targeted). A large database should permit to address two issues: How can the individual classification-power of various features be compared (cf. [6], [30]) ? How should supervised classification strategies (relying on sets of selected features) be implemented and compared ?

REFERENCES

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

[2] G. A. Macones, G. D. Hankins, C. Y. Spong, J. Hauth, and T. Moore, "The 2008 national institute of child health and human development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines," *Obstet. Gynecol.*, vol. 112, no. 3, pp. 661–666, 2008.

[3] FIGO, "Guidelines for the use of fetal monitoring," *Int. J. Gynaecol. Obstet.*, vol. 25, pp. 159–167, 1987.

[4] I. Amer-Wahlin, C. Hellsten, H. Hagberg, A. Herbst, I. Kjellmer, H. Lilja, C. Mansson, L. Martensson, P. Olofsson, A. Sundstrom, and K. Marsál, "Cardiotochography only versus cardiotochography plus st analysis of fetal electrocardiogram for intrapartum fetal monitoring: a swedish randomised controlled," *Trial. Lancet.*, vol. 358, no. 9281, pp. 534–538, 2001.

[5] W. Jarish and J. S. Detwiler, "Statistical modeling of fetal heart rate variability," *IEEE Trans. on Biomedical Engineering*, vol. 27, no. 10, pp. 582–589, 1980.

[6] B. Fulcher, A. Georgieva, C. Redman, and N. Jones, "Highly comparative fetal heart rate analysis," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 28 2012-sept. 1 2012, pp. 3135 –3138.

[7] J. Bernardes and D. Ayres-De-Campos, "The persistent challenge of fetal heart rate monitoring," *Current Opinion in Obstetrics and Gynecology*, vol. 22, no. 2, pp. 104–109, 2010.

[8] M. J. Stout and A. G. Cahill, "Electronic fetal monitoring: past, present, and future." *Clin Perinatol*, vol. 38, no. 1, pp. 127–42, vii, Mar 2011. [Online]. Available: http://dx.doi.org/10.1016/j.clp.2010.12.002

[9] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *Science*, vol. 213, no. 4504, pp. 220–222, 1981.

[10] J. V. Laar, M. Porath, C. Peters, and S. Oei, "Spectral analysis of the fetal heart rate variability for fetal surveillance: review of the literature," *Acta Obstetrica et Gynecologica*, vol. 87, pp. 300–306, 2008.

[11] E. Salamalekis, P. Thomopoulos, D. Giannaris, I. Salloum, G. Vasios, A. Prentza, and D. Koutsouris, "Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks," *BJOG*, vol. 109, no. 10, pp. 1137–1142, Oct 2002.

[12] H. Helgason, P. Abry, P. Goncalvès, C. Gharib, P. Gaucherand, and M. Doret, "Adaptive multiscale complexity analysis of fetal heart rate," *IEEE Trans. on Biomedical Engineering*, vol. 58, no. 8, pp. 2186–2193, 2011.

[13] M. Ferrario, M. G. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress," *IEEE Trans Biomed Eng*, vol. 53, no. 1, pp. 119–125, 2006.

[14] P. Hopkins, N. Outram, N. Lofgren, E. Ifeachor, and K. Rosen, "A comparative study of fetal heart rate variability analysis techniques," in *Proc. of the 28th IEEE EMBS Conf.*, 2006, pp. 1784–1787.

[15] M. Cost, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 71, no. 2 Pt 1, p. 021906, Feb 2005.

[16] R. Barbieri, E. C. Matten, A. A. Alabi, and B. E. M., "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *Am. J. Physiol Heart Circ. Physiol.*, vol. 288, p. H424?H435, 2005.

[17] P. Abry, S. Roux, V. Chudáček, P. Borgnat, P. Goncalves, and M. Doret, "Hurst exponent and intrapartum fetal heart rate: Impact of decelerations," in *The 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2013)*, 2013.

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

[19] 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, pp. 259–266, 2011.

[20] R. F. Leonarduzzi, G. Schlotthauer, and M. E. Torres, "Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia," in *Proc. 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010)*, 2010, pp. 110 –113.

[21] S. Mallat, "Group invariant scattering," *Pure and Applied Mathematics*, vol. 10, no. 65, pp. 1331–1398, 2012.

[22] J. Anden and S. Mallat, "Multiscale scattering for audio classification," in *Proc. ISMIR*, vol. 103, New-York, 2011.

[23] J. Bruna, "Intermittent process analysis with scattering moments," *PhD Thesis*, 2012.

[24] V. Chudáček, J. Anden, S. Mallat, P. Abry, and M. Doret, "Scattering transform for intrapartum fetal heart rate characterization and acidosis detection," in *35th Annual International Conference of the IEEE EMBS*, 2013, pp. 2898–2901.

[25] J. Bruna, S. Mallat, E. Bacry, and J. Muzy, "Intermittent process analysis with scattering moments," *arXiv:1311.4104*, 2013.

[26] J. V. Laar, M. Porath, C. Peters, and S. Oei, "Spectral analysis of the fetal heart rate variability for fetal surveillance: review of the literature," *Acta Obstetrica et Gynecologica*, vol. 87, pp. 300–306, 2008.

[27] P. Abry, H. Wendt, S. Jaffard, H. Helgason, P. Gonçalvès, E. Pereira, C. Gharib, P. Gaucherand, and M. Doret, "Methodology for multifractal analysis of heart rate variability: From LF/HF ratio to wavelet leaders," in *Proceeding of the IEEE Engineering in Medicine and Biology Conference. (IEEE EMBS), Buenos Aires, Argentina*, 2010.

[28] G. Hripcsak and A. S. Rothschild, "Agreement, the f-measure, and reliability in information retrieval," *Journal of the American Medical Informatics Association*, vol. 12, no. 3, pp. 296–298, 2005. [Online]. Available: http://jamia.bmjjournals.com/content/12/3/296.short

[29] B. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta*, vol. 405, pp. 442–451, 1975.

[30] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: the empirical structure of time series and their methods," *Journal of The Royal Society Interface*, vol. 10, no. 83, 2013. [Online]. Available: http://rsif.royalsocietypublishing.org/content/10/83/20130048.short