

# Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity

Nicolas Brodu <nicolas.brodu@univ-rennes1.fr>\*;  
 Fabien Lotte <fprlotte@i2r.a-star.edu.sg>†  
 Anatole Lécuyer <anatole.lecuyer@irisa.fr>‡

February 2010

## Abstract

In this paper, we introduce two new features for the design of electroencephalography (EEG) based Brain-Computer Interfaces (BCI): one feature based on multifractal cumulants, and one feature based on the predictive complexity of the EEG time series. The multifractal cumulants feature measures the signal regularity, while the predictive complexity measures the difficulty to predict the future of the signal based on its past, hence a degree of how complex it is. We have conducted an evaluation of the performance of these two novel features on EEG data corresponding to motor-imagery. We also compared them to the most successful features used in the BCI field, namely the Band-Power features. We evaluated these three kinds of features and their combinations on several real EEG signals. Results obtained with various EEG datasets show that our novel features can lead to BCI designs with improved classification performance, notably when using and combining the three kinds of feature (band power, multifractal cumulants, predictive complexity) together.

## 1 Introduction

Brain-Computer Interfaces (BCI) are communication systems that enable users to send commands to a computer by using only their brain activity [29], this activity being generally measured using ElectroEncephaloGraphy (EEG). Most EEG-based BCI are designed around a pattern recognition approach: In a first step features describing the relevant information embedded in the EEG signals are extracted [5]. They are then feed into a classifier which identifies the class of the mental state from these features [19]. Therefore, the efficiency of a BCI, in terms of recognition rate, depends mostly on the choice of appropriate features and classifiers. Despite the large number of features that have been explored to design BCI [5], the performances of current EEG-based BCI are still not satisfactory, and the BCI community has stressed the need to further explore and design alternative features [21].

In this paper, we focus on feature extraction from EEG signals for the design of BCI based on motor imagery (MI) [24]. MI corresponds to the imagination of limb movements (e.g., hand or feet) without actual output. It is known to trigger brain signals variations in specific frequency bands that can be used to drive a BCI [24]. The contribution of this paper is two-fold. First, we introduce two new features for MI classification in EEG-based BCI. These two new features are based on: 1. multifractal cumulants [28] and 2. predictive complexity [10]. Second, we perform systematic comparisons and analysis of the performance of these two new features, together with the most successful feature for motor-imagery based-BCI, namely, band power feature [15].

The first new feature, namely, multifractal cumulants, can be seen as a statistic on inter-frequency band relations. This is particularly relevant for BCI as this information is generally ignored in current motor imagery-based BCI designs, mostly based on the sole power in different frequency bands. It should be mentioned that a preliminary study of this kind of feature has been presented in a conference paper [9]. The second new feature is based on the statistical complexity and predictive properties of the time

---

\*Nicolas is with the Laboratoire du Traitement du Signal et de l'Image (LTSI) at University of Rennes 1, Campus de Beaulieu, Bât 22, 35042 Rennes Cedex - France.

†Fabien is with the Institute for Infocomm Research (I2R), Signal Processing Department, Brain-Computer Interface Laboratory, at 1 Fusionopolis way, 138632, Singapore.

‡Anatole is with the Institut National de Recherche en Informatique et en Automatique (INRIA), at Campus universitaire de Beaulieu, Avenue du Général Leclerc, 35042 Rennes Cedex - France.

series [10]. The information (quantified in bits) that is extracted this way measures how difficult it is to make an optimal prediction based on past information. It is null both for totally ordered and totally random systems, and increases in between. It has been applied already to single simulated neurons [8] and a related form was applied to measure synchronisation in the brain [17]. The assumption is that performing a mental task (e.g., motor imagery) makes the EEG signal either more or less predictable, which can be detected by a classifier when quantified by this second new feature.

Concerning the study of band power features for BCI design, we compare different techniques to extract such features and their merits, hence complementing the study performed in [15] and giving another interpretation of the results. In order to gather knowledge about the features considered in this paper, i.e., multifractal cumulants, predictive complexity and band power, all the evaluations and comparisons are performed using a variety of machine learning classifiers, on several real EEG data sets. Moreover, these features are not only evaluated on their own, but also in combination using different methods, as recommended in [13]. Our results suggest that our two novel features can improve the performances of current BCI systems when combined with band power features.

The remainder of this paper is organised as follows: Sections 2 and 3 present the two new features that we propose, while Section 4 describes different methods for estimating the power information from the signal. Section 5 presents the experiments, including the data sets that were used. The results are then discussed. Section 6 concludes this study.

The code used for all the experiments in this paper is provided as Free/Libre software. The data used is available online and all the presented results are reproducible independently.

## 2 Predictive complexity measure

This paper introduces for the first time the predictive complexity measure of [10] in the context of EEG classification.

The intuitive idea behind this feature is to quantify the amount of information that is necessary to retain from the past of the series in order to be able to predict optimally the future of the series [26], and using the sum of square error on the predicted series for the optimality criterion [10].

We had indication from related previous works that the feature could be relevant for EEGs: – At the level of a single neuron: Statistical complexity decreases while learning, in an artificial spiking neural network [8]. – At the level of the brain: Information coherence and synchronisation mechanism between communities of neurons are presented in [17], relying on related techniques.

### 2.1 Decisional states and the corresponding complexity measure

Informally, the idea behind the decisional states is to construct a Markovian automaton [12, 26] whose states correspond to taking the same decisions [10], according to a user-defined utility function.

The complexity of the series is then computed as the mutual information between the internal states of the Markovian automaton, and the series itself. The complexity is null for a very regular series, for example a constant series or a series where we always take the same decision: There is only a single state in the automaton. Similarly the complexity of a completely random series is also null: it can be modelled by successive independent draws from a fixed probability distribution, whose expected utility we take to make our decision. This leads again to a single Markovian automaton state, hence a null complexity. The complexity measure increases only for more complicated series with many internal states (i.e. many distinct probability distributions of what happens next, depending on what previously happened, leading to different decisions).

Presumably when the EEG corresponds or not to some functional activity, the complexity of the series should change. The idea is to plug machine learning techniques for monitoring that change.

#### Formal description

Formally, let  $(s_t)$  be a time series, with  $t$  the time index. Let  $s_t^{-\infty} = (s_u)_{u \leq t}$  and  $s_t^{+\infty} = (s_u)_{t < u}$  be the past and future histories at time  $t$ . In practice and for real measures, the time range is finite:  $0 \leq t \leq T$ . Similarly we measure the past and future histories with finite horizons:  $s_t^{-h} = (s_u)_{t-h \leq u \leq t}$  and  $s_t^{+k} = (s_u)_{t < u \leq t+k}$ .

Let us now consider predicting the future from the past statistically: we seek to determine  $P(s_t^{+k} | s_t^{-h})$  at each time  $t$  (possibly with  $h$  or  $k$  infinite). Assuming the system is conditionally stationary, that distribution does not change through time: the same causes produce the same consequences. Let then

$S^{-h} = \{s_t^{-h}\}_{\forall t}$  and  $S^{+k} = \{s_t^{+k}\}_{\forall t}$  the sets of all past and future histories. We will drop the time indices from now on to indicate the time shift invariance.

The causal states  $\zeta$  are defined as the equivalence classes of past histories with the same conditional distribution of futures:  $\zeta(s^{-h}) = \{x \in S^{-h} : P(s^{+k}|x) = P(s^{+k}|s^{-h})\} = \{x : x \stackrel{c}{\equiv} s^{-h}\}$ . Knowing the causal state at the current time is the minimal information needed for making optimal predictions [26] using the full conditional probability distribution.

In the discrete case the series are strings of symbols drawn from an alphabet  $\mathcal{A}$ . Each time step implies a symbol transition, which possibly leads to a different causal state. The corresponding automaton is called the  $\epsilon$ -machine [12].

Let us now introduce a utility function  $u : (S^{+k})^2 \mapsto \mathbb{R}$ , such that  $u(r^{+k}, s^{+k})$  quantifies the gain (positive) or loss (negative) when the user relied on the prediction  $r^{+k}$  while  $s^{+k}$  actually happened. We can now define an expected utility:  $\mathbb{U}[r^{+k}|s^{-h}] = \mathbb{E}_{s^{+k} \in S^{+k}} [u(r^{+k}, s^{+k})|s^{-h}]$ , quantifying what utility can be expected on average when choosing the prediction  $r^{+k}$  for the current system state  $s^{-h}$ . The set of optimal predictions, realising the maximal expected utility, can now be defined as  $Y(s^{-h}) = \operatorname{argmax}_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k}|s^{-h}]$ .

The following equivalence relations  $\stackrel{p}{\equiv}$ ,  $\stackrel{u}{\equiv}$ , and  $\stackrel{d}{\equiv}$  naturally extend the causal state equivalence relation  $\stackrel{c}{\equiv}$ , taking into account the utility function:

- $r^{-h} \stackrel{p}{\equiv} s^{-h}$  when  $Y(r^{-h}) = Y(s^{-h})$ , with the corresponding *iso-prediction sets* as equivalence classes. All past histories within the same class lead to choosing the same predictions, even though the expected utility may change from one past history to the other.
- $r^{-h} \stackrel{u}{\equiv} s^{-h}$  when  $\max_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k}|r^{-h}] = \max_{r^{+k} \in S^{+k}} \mathbb{U}[r^{+k}|s^{-h}]$ , with the corresponding *iso-utility sets* as equivalence classes. All past histories within the same class lead to the same maximum expected utility, even though the optimal predictions to choose for reaching this utility are not specified.
- $r^{-h} \stackrel{d}{\equiv} s^{-h}$  when  $r^{-h} \stackrel{p}{\equiv} s^{-h}$  and  $r^{-h} \stackrel{u}{\equiv} s^{-h}$ . The intersection of the above iso-utility and iso-prediction sets define the *decisional states*:  $\Psi(s^{-h}) = \{x \in S^{-h} : x \stackrel{p}{\equiv} s^{-h}, x \stackrel{u}{\equiv} s^{-h}\}$ . We assume that when both the maximal expected utility and the optimal predictions are the same, the user will reach the same decisions. In other words, the utility function encodes all the user needs to know to reach a decision.

It can be easily shown [10] that the causal states sub-partition the decisional states. That is to say, the causal states have lost their minimality property due to the fact we are not interested in the full conditional distribution of futures but only in the optimal decisions with respect to a user-defined utility function.

The causal states are an intrinsic property of the data set. The mutual information  $M_c = I(s^{-h}; \zeta(s^{-h}))$  between the causal  $\zeta(s^{-h})$  state and the series of  $s^{-h}$  defines an intrinsic measure, the statistical complexity [12], quantifying how hard it is to get the conditional distribution of futures from the current observed past.

The decisional states correspond to the structure implied by the user utility function on top of the causal states. As for the causal states, knowing the decisional state at the current time is the information needed for making optimal predictions maximising the user-defined utility function. The mutual information  $M_d = I(s^{-h}; \Psi(s^{-h}))$  similarly defines a complexity measure for how hard it is to make these predictions, called decisional complexity by analogy with the statistical complexity.

## 2.2 Application to EEG data

In the present study the chosen utility function is the negative sum of square error:  $u(r^{+k}, s^{+k}) = -\|r^{+k} - s^{+k}\|^2$ .

Each EEG series is split in time windows of  $h + k$  samples, defining each a  $(s^{-h}, s^{+k})$  pair. These observations are fed in the reconstruction algorithm presented in [10]. That algorithm estimates the joint probabilities  $p(s^{-h}, s^{+k})$  using a kernel density estimation with Gaussian kernels. The conditional probabilities are then computed by integration over  $S^{+k}$ :  $p(s^{+k}|s^{-h}) = p(s^{-h}, s^{+k}) / \int_{\sigma^{+k} \in S^{+k}} p(s^{-h}, \sigma^{+k})$ . These are then clustered into causal state estimates  $\zeta(s^{-h}) = \{x \in S^{-h} : p(s^{+k}|x) = p(s^{+k}|s^{-h})\}$  using connected components up to a fixed Bhattacharyya distance threshold (chosen to be 0.05) between

the conditional distributions. The utility function is applied on top of the aggregated causal states distributions  $p(s^{+k}|\zeta) = \text{avg}_{s^{-h} \in \zeta} p(s^{+k}|s^{-h})$  in order to get the expected utilities  $\mathbb{U}[r^{+k}|\zeta]$ . Finally, the causal states are themselves clustered into iso-utility and iso-prediction sets, which are intersected to get the decisional states  $\omega$ .

The decisional complexity is the feature that is extracted from the EEG series. We thus get one feature per electrode.

### 3 Multifractal cumulants

The multifractal formalism is described in details in [28, 23]. This section presents a short overview for the needs of this document.

Intuitively, the multifractal cumulants of the signal capture a signature of inter-band relations (see below). This contrasts to the usual power in each frequency band defined in the next section. As shown in [9] the multifractal spectrum can in itself be used for EEG classification. When considering multifractal in addition to power band feature vectors, the resulting combination may improve the classification accuracy.

The method we chose for extracting the multifractal spectrum is a discrete wavelet transform of the signal, out of which we extract the wavelet leader coefficients [2]. Following the directions of [28] we then use the cumulants of the leaders as the features for classification, unlike what we previously did in [9].

Let  $x(t)$  be the signal to analyse. One view on multifractal analysis [20] is to relate the statistical properties of  $x(t)$  and of a scaled version of it  $x(at)$ . In terms of frequency analysis, that scaling in time corresponds to a frequency shift. Hence, another view of the multifractal cumulants feature is that they characterise some form of inter-frequency information, as mentioned in the introduction of this section.

More precisely:

- The signal  $x(t)$  is decomposed using a Discrete Wavelet Transform to get the wavelet coefficients  $w(s, t_s)$  at each dyadic scale  $s$  and time interval  $t_s$ . This decomposition splices the time-frequency plane as depicted in Fig. 1.
- The wavelet leaders at each scale  $s$  are then extracted by computing the maxima of the wavelet coefficients over all samples involved for computing  $w(s, t_s - 1)$ ,  $w(s, t_s)$  and  $w(s, t_s + 1)$  (including lower scales) [2].
- Instead of performing a Legendre transform, or a direct Holder exponent density estimation as in [9], we use here the recent technique introduced by [28] and compute the wavelet leader cumulants of order 1 to 5. As noted in [28] the first few cumulants already contain most of the information useful in practice for characterising the distribution of the Holder exponents. For a classification task this information can now be exploited in a more condensed form.
- The 5 first cumulants are computed for the leaders at each scale  $s$ . Considering there is at most  $L$  levels of wavelet transform in a signal of size between  $2^L$  and  $2^{L+1}$ , we get a total of  $5 \times L$  cumulants for the signal, that progressively encompass more and more frequency bands as the scale increases. These  $5 \times L$  cumulants are used as the feature vector.

This method can be quite sensitive to the presence of electromagnetic interferences at 50 Hz. We thus pre-filter the signals as described in section 5.1 before proceeding to the multifractal cumulants estimation.

### 4 Power in frequency bands

Time-frequency decomposition is in itself a wide topic of research, which has been extensively studied [14]. A thorough comparison of different power feature extraction techniques was presented in [15], together with their use in classification of motor imagery. This section aims at complementing and reproducing independently the observations in that survey. Notable difference between our approach and [15] include:

- In [15] the spectrum estimation in the different bands are added in a single feature. We aim here at maintaining the separation of relevant frequency bands for the classifiers. In [15] the power density is extracted along each time-frequency decomposition atom (See Fig 1), and the energy is summed over frequency components at that time. We instead prioritise the frequency representation and keep individual frequency bands, while summing along the time axis. This is justified physiologically by considering that cerebral activity is more stable in the frequency than in the time domain (i.e. we are looking for frequency signatures rather than time patterns).

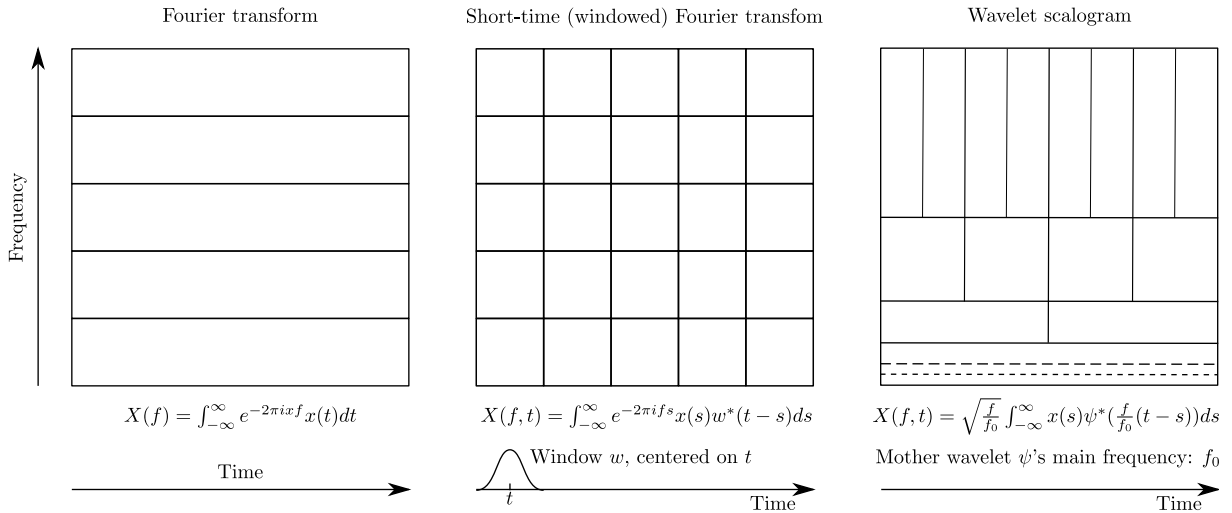


Figure 1: Splicing the time-frequency plane

- We use the log-transform of the energy in each frequency band as the discriminant feature. We decided to compute the energy on 1 Hz bands from 4 to 30 Hz, for both the C3 and C4 electrodes. One can recover the physiological  $\mu$  and  $\beta$  bands (around 8-12 Hz and 16-24 Hz) by adding the corresponding 1Hz bands to reduce the number of features. Yet we prefer to let the classifiers exploit all available information, with possible variations between individuals, rather than fixing the  $\mu$  and  $\beta$  bands a priori.

The next subsections describe which features are how they extracted in the present study, with comparisons to [15] when applicable. We considered using the spectrogram, the Wigner-Ville distribution, the Morlet wavelet scalogram, the periodogram on the whole signal, direct band-pass filtering using a Butterworth filter, and power spectral density estimation using an auto-regressive model.

#### 4.1 The spectrogram method

A usual way to extract the power in frequency bands is to use the spectrogram method [14] to get a time-frequency signal decomposition:

- The signal is convolved with a windowing function that is localised in time around  $t$ . A Fourier transform is then performed on the convolved signal, non null only on a sliding chunk centred on  $t$ , leading to a time-frequency representation  $X(f, t)$  of the original signal  $x(t)$  (see Fig 1, second case).
- The energy marginals in the frequency domain [4] allows us to estimate a spectrum:  $E_f = \sum_t |X(f, t)|^2$
- The energy in each band  $B$  is obtained by summing the marginals in that band:  $E_B = \sum_{f \in B} E_f$ .

As explained in the introduction to this section, a feature vector is then formed from the logarithms of the estimated energy in each 1Hz band.

#### 4.2 The Wigner-Ville distribution

As shown in [14] there is a compromise to make between increased precision for the energy estimation in each decomposition region of the time-frequency plane (see Fig 1), and the occurrence of artifacts. The spectrogram introduces little artifacts, and the Wigner-Ville distribution realises the opposite compromise: improved accuracy over each decomposition atom, at the cost of more artifacts.

An hypothesis is that the improved localisation in frequency could compensate the presence of the artifacts for some classifiers. We thus considered the Wigner-Ville distribution in order to compare our results (and our different choice of feature vectors) with [15].

### 4.3 The Morlet wavelet scalogram

The scalogram realises the third slicing of the time-frequency plane represented on Fig 1. A mother wavelet  $\psi(x)$  function is scaled and shifted and forms a basis for the decomposition. The frequency interpretation comes from the fact that if the mother wavelet is localised at a given frequency  $f_0$ , then a scaled version  $\psi(ax)$  is localised at frequency  $\frac{f_0}{a}$  by the scaling property of the Fourier transform. The wavelet decomposition formula can thus be rewritten as we did in Fig. 1 to highlight  $\frac{f}{f_0}$  as a scaling factor.

In [15] the mother wavelet is carefully chosen so the relevant  $\mu$  and  $\beta$  bands are well covered in the time-frequency plane, since they combine the energy contributions in these bands into a single feature. In the present study we use the facility provided by [4] to adapt the wavelet so that  $f = \frac{f_0}{a}$  ranges between 4 and 30 Hz: For a Morlet wavelet  $f_0$  can be chosen arbitrarily, and the scale parameter  $a$  is adapted to maintain the selected frequency range. The influence of this free parameter however is to shape the effective resolutions on both axis in the time-frequency plane as shown in 1. In the present study we use a 10-fold cross-validation on the training set in order to select the best time-frequency resolution compromise.

### 4.4 Periodogram on the whole signal

The spectrogram method considered sliding chunks because it is primarily used for time-frequency decomposition: the energy is localised with a fixed resolution in both the frequency and time domain. However here we only perform a classification step at the end of trials. Thus we are only interested in the frequency components over a fixed signal duration, not at each instant. Applying an FFT transform to the whole signal and proceeding directly to the power estimation better matches the structure of the time-frequency decomposition we need, as depicted in Fig. 1.

Unfortunately as the signal is usually finite and non-periodic this technique introduces spurious artifacts in otherwise unrelated frequency bands (spectral bias). A workaround is to apply a window function on the whole signal, but we loose some information as the window would smoothly damps it to 0 on each end. The spectrogram copes with this by sliding each chunk in time, thus getting a correct estimate over each chunk centre.

A second workaround is to remove the 0 and low-frequency components so as to detrend the signal: This operation centres the signal around 0, and the periodisation implicitly operated by the FFT results in less artifacts. We applied this second workaround with the linear filters presented in Section 5.1: at least 50dB attenuation at 0Hz and less than 1dB change at 4Hz.

The periodogram is then simply computed:

- The signal  $x(t)$  is taken as a whole, without applying a window function, but filtered as aforementioned.
- A Discrete Fourier Transform is applied, giving a frequency representation  $X(f)$  of the whole signal, including artifacts due to the non-periodic nature of the signal (attenuated by the FIR pre-filtering).
- Power is extracted in narrow bands  $B$  (of 1 Hz in this example):  $E_B(X) = \sum_{f \in B} |X(f)|^2$ .

### 4.5 Band-pass filtering

It is also possible to directly apply a band-pass filter to the signal  $x(t)$  in order to damp out frequencies outside each selected band  $B$ : the resulting filtered signal  $y(t)$  presumably contains only the selected frequencies. Assuming a perfect block filter, it is then possible to estimate the energy in these bands by squaring the magnitude of the time series samples directly:  $E_{band} = \sum_t |y(t)|^2$ .

This is a frequent feature found in the BCI field [24]. This is the method that is performed by default in the OpenViBE software [25], with a Butterworth IIR filter of order 4, and that was also used for [9].

In the present study we select the order of the Butterworth filter on the training data set in a 10-fold cross-validation step using a linear discriminant classifier.

### 4.6 Power spectral density estimation using an auto-regressive model

The Yule-Walker algorithm can be used to train an auto-regressive model. The power spectral density is then derived from the model's response. As in [15] we estimate the order of the auto-regressive model with a 10-fold cross-validation on the training data set (using linear discriminant analysis). However the

feature we extract is the log-transform of the energy in each of the bands. We did not consider using directly the AR model coefficients as the feature vector, as this was covered already in [15].

## 5 Evaluation

### 5.1 Data sets

Three data sets, for a total of four subjects, were used for evaluation in this study. Data for three subjects come from the international BCI competitions<sup>1</sup> [6][7]. The data for the fourth subject was acquired at INRIA (French National Research Institute on Computer Science and Control) Rennes-Bretagne Atlantique, using the OpenViBE software platform [25]<sup>2</sup>.

#### 5.1.1 BCI competition II, data set III (1 subject)

This data set was captured at the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [1]. The data contains 280 trials sampled at 128 Hz. During each trial, the subject is presented with a visual cue indicating either left or right at random, and shall then imagine a movement of the corresponding hand during 6s. There was 140 trials of left class (left hand motor imagery) and 140 trials of the right class (right hand motor imagery). EEG were recorded using the C3, C4 and Cz electrodes, however, for the purpose of this evaluation, we used only the C3 and C4 electrodes as recommended in [18]. More details about this data set can be found in [6].

The data was already preprocessed by a band-pass filter between 0.5 and 30 Hz. Unfortunately, the nature of the filter is not specified, and the DC component was not well removed. Since this interferes in particular with the Periodogram power density estimation method, as well as with the algorithm for estimating the series complexity, a new filter is applied with the following characteristics:

- FIR filter obtained using METEOR [27]
- Less than 1dB change in 4-30Hz with a linear phase response
- At least -50dB at 0 Hz (and above 40 Hz, though in the present case the signal is already filtered)
- 1/4 sec delay

We then used all the available filtered data over the feedback period in order to extract the features.

#### 5.1.2 BCI competition III, data set IIIb (2 subjects)

This data set was also captured at the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [1]. It originally consists of EEG signals from 3 subjects who performed left hand and right hand motor imagery. However, for the purpose of this study, only subjects labeled S4 and X11 were used. Indeed, EEG signals for subject O3VR were recorded using a different protocol and the data file provided online contained erroneously duplicate signals<sup>3</sup>. The experimental protocol for subjects S4 and X11 is similar to the one used for the BCI competition II data set III. The data was captured at a 125 Hz sampling rate using electrodes C3 and C4. More details about this data set can be found in [7]. EEG signals in this data set were already band-pass filtered in 0.5-30 Hz. As for the previous data set, an additional FIR filter with the same characteristics as the previous one was applied, for the same reasons. We also used all the available filtered data over the feedback period in order to extract the features.

#### 5.1.3 OpenViBE / INRIA data (1 subject)

This data set was captured at INRIA Rennes-Bretagne Atlantique using the home-made OpenViBE free and open-source software [25]. This data set comprise EEG signals from one subject who performed left hand and right hand motor imagery. 560 trials of motor imagery (280 trials per class) were recorded over a 2 week period. Data were collected using the same experimental protocol as the one used for the BCI competition data, i.e., the Graz BCI protocol [24]. Half of the trials (randomly selected from all experiments over time) is used for training, the remaining half for testing.

---

<sup>1</sup><http://www.bbci.de/competition/>

<sup>2</sup><http://openvibe.inria.fr/?q=datasets>

<sup>3</sup>See [http://www.bbci.de/competition/iii/desc\\_IIIb\\_ps.html](http://www.bbci.de/competition/iii/desc_IIIb_ps.html) for details

EEG data was sampled at 512 Hz and recorded using the following electrodes: C3, C4, FC3, FC4, C5, C1, C2, C6, CP3, CP4, with a nose reference electrode. The use of such electrodes enables us to apply a discrete Laplacian spatial filter [16] over C3 and C4 in order to obtain better signals, as recommended in [22].

The data was preprocessed by a FIR filter (designed with METEOR [27]) with the following characteristics:

- Less than 1dB change in 4-30Hz, linear phase response in this range.
- At least -50dB at 0 Hz and above 50 Hz.
- Null responses at 50 Hz and all harmonics.
- 1/4 sec delay.

We then used all the available filtered data over the feedback period in order to extract the features.

## 5.2 Comparison of the power extraction techniques for classifying real signals

The techniques presented in Section 4 are compared on the data sets. The goal is to complement and reproduce independently results from [15]. Selection of a good technique for extracting the power in selected frequency bands also helps in better asserting the contribution of other features (multifractal cumulants, predictive complexity) when considered in combination to the power density feature.

Subjects → Methods ↓	BCI 2		BCI 3, S4		BCI 3, X11		OpenViBE	
	CV	Test	CV	Test	CV	test	CV	Test
Periodogram	77.9	74.3	80.9	80.4	77.8	77.4	<b>91.8</b>	<b>92.1</b>
Butterworth	73.6	72.9	81.1	80.7	78.1	75.9	88.9	91.8
Yule-Walker	80.7	76.4	<b>84.3</b>	81.3	79.8	76.1	87.1	84.3
Spectrogram	70.7	<b>82.1</b>	81.9	80.7	79.3	<b>79.3</b>	87.9	89.6
Wigner-Ville	75.7	76.4	81.5	<b>81.5</b>	76.9	<b>77.2</b>	86.8	87.5
Morlet Wavelet	80.0	74.3	<b>84.3</b>	79.3	80.7	78.5	89.3	90.4

Table 1: Classification accuracy (%) of power extraction techniques on real data sets

Table 1 shows the result of classifying the signal using each of the techniques presented in Section 4 with Linear Discriminant Analysis. The parameters described in each respective method subsection (ex: IIR filter length, AR model order, etc) were determined by searching the parameter space, and optimisation of the 10-fold cross-validation accuracy on the training set. Table 1 shows the result of the 10-fold cross-validation for the selected best parameters, for each method and each data set. The accuracy on the test set is also presented.

Unlike what is presented in [15] we do not see a clear dominance of one method over the others. Five out of the six retained methods lead to the best accuracy in at least one data set, either during cross-validation or testing. The Butterworth IIR method (the remaining one) gives results on the same order, close to the maximum for the S4 and OpenViBE subjects. We therefore cannot conclude that this feature is worse than the others based only on such a small sample size: this IIR band-pass filtering may very well lead to the best accuracy for other data sets.

Yet in each experiments the variations from one method to the other are quite important: we observe a variation between 2 to 8 % between the best and worse power estimation technique.

As a conclusion, there is no one-fit-all preferred method: Several power estimation techniques shall be tried in a first off-line phase for a given data set, and one of them shall be selected for the real-time online classification (or possibly a combination of them). We emphasise however that one shall not rely on a single technique, as the best power in frequency bands extraction method changes from one individual to another in this study.

## 5.3 Results obtained with all features and their combinations

The goal of this section is to show what results can be obtained with each feature considered in isolation, and the effect of combining them. The hypothesis we wish to test is that each feature extracts a different information from the signal, and that combining them can improve the classification accuracy.

### 5.3.1 Features

Table 1 also shows that 10-fold cross-validation is a poor estimator for what is the best technique to use on the Test set. Despite this known issue, we still use the cross-validation performance so as to choose which power density estimation technique to choose in the next experiments. This is justified by the fact that for real on-line signals, we do not have access to the test set, and that the cross-validation results are then a reasonable option to choose.

For the experiments below, the following methodology was applied:

- The power in frequency bands feature was extracted using the method giving the best cross-validation performance on the training set. All bands were kept between 4 and 30 hertz, for each electrode: this leads to a 52-dimensional feature vector.
- The multifractal cumulants feature was extracted according to the method described in Section 3. Cross-validation on the training set was used to determine the following properties: The wavelet support (a Daubechies wavelet with 6 coefficients was selected for each data set), and the number of decomposition levels to retain (3 for the BCI II subject, 5 for the others). Since we have 5 cumulants per level per electrode, this leads to a 30-dimensional vector for the BCI II subject, and a 50-dimensional vector for the others.
- Parameters for estimating the predictive complexity feature were also determined by maximising the cross-validation performance on the training set. The number of points to retain from the past and the future were determined, as well as a sub-sampling factor for the series, and the kernel size used for estimating the conditional probability distributions, and whether to work on the raw series or the first differences. Cross-validation selected 1 point in the future, 5 points from the past with a sub-sampling factor of 8 for the OpenViBE subject, and 6 points from the past with a sub-sampling factor of 2 for the others. These sub-sampling parameters correspond to having a sampling frequency of 62.5 Hz for the S4 and X11 subjects and 64 Hz for the others, which thanks to Nyquist theorem matches the filtering operation that was performed on the signals: cross-validation selected the most economical sub-sampling parameter that still captures the remaining frequencies in the signal between 4 and 30 Hz. All sub-sampled signals (ex: all 8 possible series for a sub-sampling by factor 8) were kept for building the statistics in the complexity computations.

Therefore the feature vectors include:

- 52 power features for frequencies between 4 and 30 Hz. (25 for each C3 / C4 electrode)
- 30, or 50 features for multifractal cumulants.
- 2 features for the predictive complexity (one for each electrode)

As a side remark we shall point out that cross-validating all the above parameters required a quite consequent computational power, which was obtained using the BOINC distributed computing infrastructure [3]. Nevertheless, once the optimal parameters were determined off-line by cross-validation, the computational requirements for extracting the features are nothing exceptional: the features can be computed in real time on a standard PC.

### 5.3.2 Classifiers and combination rules

Each feature implies a different representation of the data. But some classifiers might be more sensitive than others on the shape of that space. For example, a nearest neighbours classifier is more sensitive to the distance between individual data points in the feature space than a linear classifier which splits that space by a hyperplane. We have thus selected classifiers that take different stance on the data, detailed below, in order to assert the value of each feature.

In this paper, we tested three different features, three combinations of two of them, and one combination of the three of them altogether. Some classifiers like Support Vector Machines and Multi-Layer Perceptrons have internal parameters that are set by optimising the 10-fold cross-validation performance on the training set. Then the experiment was repeated using 30 different random seeds in order to build an estimate of the variance in their results. The number of calculations to perform is  $(3 + 4N_{Comb}) * (N_{ClassifD} + 30 * N_{ClassifCV} * N_{CV}) * N_{Subjects}$  where  $N_{Comb}$  is the number of combination rules,  $N_{ClassifD}$  is the number of deterministic classifiers,  $N_{ClassifR}$  is the number of classifiers depending on the random seed, and  $N_{CV}$  is the number of cross-validation steps to perform.

Previous work experiments [9] have shown that some classifiers and combination rules perform better than others. Given the number of tests to perform in order to compute all comparisons, the following classifiers and combination rules were selected amongst possible classifiers commonly found in the BCI literature [19].

- Linear classifiers:
  - The hyperplane separating the data with Least Squared Error (abbreviated LSE). This classifier returns the predicted class for each instance without probability estimation, as 0 or 1.
  - Linear Discriminant Analysis (LDA). This classifier uses a sigmoid output for its confidence that it predicts a given instance class.
  - A Support Vector Machine with a Linear kernel, hence margin maximisation in the data space (SVML). This classifier uses the probability estimation technique of [11, 30] for its confidence in each instance classification.
- Non-linear classifiers:
  - A Support Vector Machine using a Gaussian kernel (SVMG). This classifier uses the probability estimation technique of [11, 30] for its confidence in each instance classification.
  - The classifier attributing each test point to the class of its nearest neighbour in the training set (NNE). This classifier uses an internal vote between the neighbors for the instance class and uses the proportion of votes as a classification probability.
  - A multi-layer perceptron, whose number of hidden units were determined by cross-validation as aforementioned (MLP). The network has two output nodes, one for each class, and their difference gives the confidence for classifying a given input instance.
- Combination rules:
  - Single Vector: Put all features together as a single higher dimensional vector.
  - Average: Classify with each feature separately. Then for each instance, perform an arithmetic average across classifiers of the class decision confidence (ex: probability output for the SVM), weighted by the global classifier performance evaluated by cross-validation.
  - Max Decision: Classify with each feature separately. Then for each instance, consider only the class with the maximum decision confidence across all features (ignores the classifiers global performance except to break ties). As this is done for each instance, the global accuracy over the data set can be higher than using either feature uniquely for all instances.

The BOINC [3] infrastructure was again used in this step in order to perform all these computations, noting that any single classifier and combination rule alone has a low computational requirement.

### 5.3.3 Results

The results given in Tables 2, 3, 4, 5, are the average classification performance (percentage of success) on the test set for all considered feature combination and classifiers. The values in parenthesis are the standard deviations observed for these results over 30 independent realisations of the same experiment, using distinct random seeds.

## 5.4 Analysis of the results

For all the subjects and classifiers, the multifractal cumulants and the predictive complexity features have provided more than 50% classification accuracy, confirming that some information was extracted from the signal. This, in itself, is a contribution of this paper.

We also observe that on the subject from BCI competition II, classification results obtained with the multifractal cumulants feature actually surpass those obtained with the power estimation technique selected by cross-validation in Section 5.2 (80.7% instead of 76.4% using Linear Discriminant Analysis). Multifractal cumulants alone then give results on par with the best power estimation method on the test set (82.1% for the Spectrogram method using Linear Discriminant Analysis, and between 80.7% and 83.6% for the multifractal cumulants method with various classifiers).

Classifiers → Features ↓   Combi. ↓		Linear			Non-Linear		
		LSE	LDA	SVML	SVMG	NNE	MLP
Power (Yule-Walker)		74.3	76.4	77.5 (0.2)	77.7 (0.2)	72.9	77.5 (0.2)
Multifractal Cumulants		80.7	80.7	82.3 (0.3)	81.1 (0.3)	83.6	82.5 (0.2)
Predictive Complexity		77.9	77.9	79.0 (0.3)	77.9 (0.3)	74.3	80.5 (0.2)
P,M	Average	81.4	78.6	79.2 (0.5)	78.7 (0.3)	80.7	81.4 (0.2)
	Max Decision	80.7	79.3	79.2 (0.5)	78.9 (0.3)	79.3	81.4 (0.2)
	Single vector	75.0	77.1	82.2 (0.1)	80.9 (0.1)	79.3	81.3 (0.2)
P,C	Average	78.6	77.1	75.0 (0.5)	74.1 (0.2)	77.1	82.2 (0.1)
	Max Decision	80.0	78.6	75.4 (0.4)	74.2 (0.2)	77.9	82.2 (0.1)
	Single vector	76.4	75.0	79.5 (0.1)	79.7 (0.2)	78.6	79.6 (0.2)
M,C	Average	80.7	80.0	77.9 (0.4)	76.3 (0.4)	80.7	83.6 (0.2)
	Max Decision	81.4	80.7	78.0 (0.4)	76.3 (0.4)	80.7	83.5 (0.2)
	Single vector	79.3	79.3	83.3 (0.3)	81.4 (0.3)	<b>85.0</b>	82.1 (0.2)
P,M,C	Average	81.4	80.7	77.2 (0.4)	76.7 (0.3)	79.3	82.6 (0.2)
	Max Decision	82.9	80.7	75.7 (0.5)	74.0 (0.4)	80.0	83.2 (0.2)
	Single vector	77.9	74.3	82.2 (0.1)	82.0 (0.1)	79.3	81.1 (0.2)

Table 2: Results for the subject of BCI competition II, data set III

Classifiers → Features ↓   Combi. ↓		Linear			Non-Linear		
		LSE	LDA	SVML	SVMG	NNE	MLP
Power (Yule-Walker)		81.3	81.3	82.1 (0.1)	82.8 (0.1)	76.1	81.9 (0.1)
Multifractal Cumulants		74.8	74.4	73.1 (0.2)	72.5 (0.2)	58.0	73.0 (0.1)
Predictive Complexity		65.2	65.0	64.7 (0.0)	64.7 (0.1)	61.9	64.9 (0.0)
P,M	Average	83.0	83.0	83.0 (0.1)	83.6 (0.1)	76.1	82.0 (0.1)
	Max Decision	81.7	82.6	83.1 (0.1)	<b>83.9 (0.1)</b>	75.9	81.7 (0.1)
	Single vector	80.7	80.7	80.5 (0.1)	80.9 (0.1)	73.5	80.8 (0.1)
P,C	Average	82.0	81.1	81.9 (0.1)	82.5 (0.1)	75.0	80.9 (0.1)
	Max Decision	78.9	81.3	81.3 (0.1)	81.9 (0.1)	74.3	80.6 (0.1)
	Single vector	81.9	81.9	81.6 (0.2)	82.1 (0.2)	76.5	81.6 (0.1)
M,C	Average	74.6	74.6	74.2 (0.1)	73.6 (0.1)	62.8	74.4 (0.1)
	Max Decision	75.6	75.2	74.2 (0.1)	73.7 (0.2)	63.0	74.4 (0.2)
	Single vector	74.1	73.9	73.1 (0.5)	73.7 (0.2)	56.7	73.0 (0.1)
P,M,C	Average	82.6	82.2	79.0 (0.1)	79.2 (0.1)	73.7	81.1 (0.1)
	Max Decision	79.6	82.6	82.2 (0.1)	83.0 (0.2)	73.9	80.7 (0.1)
	Single vector	80.6	80.4	81.0 (0.1)	80.8 (0.1)	73.5	80.9 (0.1)

Table 3: Results for the S4 subject, BCI competition III data set IIIb

The present experiments consider 6 classifiers and 3 data sets in common with [9]. Except for the Nearest Neighbours classifier on the S4 subject that over-fits (99.3% accuracy on the training set, not shown in the above tables), we observe a clear improvement in the results using the multifractal estimation technique over the results from [9]. Moreover the multifractal features from [9] were not able to extract useful information for the X11 subject; while using the cumulants at different wavelet scales we can now obtain 68.7% accuracy on the training set using the multifractal feature alone for that subject.

Hence, this study and the newly introduced “cumulants at different scales” technique confirm that multifractal analysis can be useful for classifying EEG signals independently of the power-density features.

The novel complexity feature also reaches good results (much higher than 50%). This feature alone achieved a decent classification rate comparatively to the other two features on the BCI competition II subject, but an inferior rate on the other subjects. There is also the possibility that we did not extract the best possible score from this feature, as we had to discretize the parameter space quite roughly for the cross-validation phase of the kernel size used for estimating the conditional probability distributions (see Section 2.2). When the complexity feature is considered in combination with the others, however, it has the potential to improve the classification rate (See the OpenViBE and the BCI2 subjects). This can be explained by the fact we are combining a different information that was not present in the other features. In fact, for each classifier, the maximum classification rate was obtained using a combination of features in most cases. This confirms that the information extracted by either the multifractal cumulants

Classifiers → Features ↓   Combi. ↓		Linear			Non-Linear		
		LSE	LDA	SVML	SVMG	NNE	MLP
Power (Morlet wavelet)		78.5	78.5	78.9 (0.1)	78.3 (0.2)	69.4	78.7 (0.1)
Multifractal Cumulants		68.7	68.7	68.4 (0.1)	68.7 (0.1)	59.1	69.3 (0.2)
Predictive Complexity		68.9	70.2	69.1 (0.1)	69.0 (0.1)	68.5	69.7 (0.1)
P,M	Average	79.3	<b>79.8</b>	70.9 (0.2)	71.3 (0.2)	70.6	77.9 (0.1)
	Max Decision	78.1	78.0	70.7 (0.2)	70.9 (0.2)	70.4	77.6 (0.1)
	Single vector	75.7	75.7	76.2 (0.2)	76.7 (0.2)	68.0	78.0 (0.2)
P,C	Average	78.5	78.7	70.9 (0.2)	72.0 (0.2)	<b>71.9</b>	77.0 (0.1)
	Max Decision	<b>79.3</b>	78.3	70.6 (0.2)	71.9 (0.2)	<b>71.9</b>	76.5 (0.1)
	Single vector	78.1	78.0	78.7 (0.1)	77.9 (0.1)	69.8	78.4 (0.1)
M,C	Average	71.5	71.3	67.0 (0.3)	67.8 (0.3)	65.9	71.9 (0.1)
	Max Decision	71.9	72.2	67.4 (0.2)	67.9 (0.2)	66.7	71.9 (0.1)
	Single vector	69.6	70.4	70.3 (0.1)	70.4 (0.1)	62.6	71.4 (0.2)
P,M,C	Average	79.1	78.0	71.6 (0.2)	70.8 (0.2)	70.7	76.4 (0.1)
	Max Decision	78.3	77.6	68.8 (0.2)	69.9 (0.2)	71.3	76.6 (0.1)
	Single vector	74.8	74.8	76.7 (0.1)	76.5 (0.1)	68.0	78.0 (0.1)

Table 4: Results for the X11 subject, BCI competition III data set IIIb

Classifiers → Features ↓   Combi. ↓		Linear			Non-Linear		
		LSE	LDA	SVML	SVMG	NNE	MLP
Power (Periodogram)		92.5	92.1	93.5 (0.2)	94.1 (0.2)	<b>84.6</b>	92.8 (0.1)
Multifractal Cumulants		85.7	85.7	86.5 (0.2)	85.8 (0.1)	72.9	86.4 (0.2)
Predictive Complexity		76.1	76.1	76.2 (0.0)	75.7 (0.4)	75.7	76.4 (0.1)
P,M	Average	92.5	<b>93.6</b>	86.4 (0.2)	86.5 (0.2)	83.6	93.2 (0.1)
	Max Decision	<b>92.9</b>	93.2	86.4 (0.2)	86.4 (0.2)	83.6	93.1 (0.1)
	Single vector	91.8	92.1	92.4 (0.2)	93.1 (0.2)	82.5	92.3 (0.1)
P,C	Average	91.8	91.4	77.6 (0.1)	78.4 (0.4)	83.2	91.9 (0.1)
	Max Decision	91.8	92.1	77.6 (0.1)	78.2 (0.4)	83.2	91.7 (0.1)
	Single vector	<b>92.9</b>	<b>92.9</b>	93.2 (0.2)	<b>94.5 (0.1)</b>	<b>84.6</b>	<b>93.7 (0.1)</b>
M,C	Average	85.7	86.1	79.1 (0.2)	78.9 (0.4)	78.6	83.4 (0.1)
	Max Decision	85.7	85.4	79.0 (0.2)	78.7 (0.4)	78.9	83.1 (0.1)
	Single vector	85.7	85.7	86.7 (0.2)	85.8 (0.2)	75.4	86.0 (0.2)
P,M,C	Average	92.1	91.4	80.2 (0.2)	80.8 (0.4)	82.5	92.0 (0.1)
	Max Decision	92.1	92.1	79.0 (0.2)	79.4 (0.5)	82.5	91.4 (0.1)
	Single vector	92.5	92.1	92.5 (0.2)	92.4 (0.2)	80.4	92.3 (0.2)

Table 5: Results for the OpenViBE data

or the predictive complexity feature complements the information contained in the power feature.

Finally, as for the selection of the method used to estimate the power in frequency bands, our analysis shows that there is no preferred way to combine the features and no preferred classifier. It is possible that some other combination method or classifier be more reliable, but given the current results we cannot recommend one particular way to combine features or classify them. However, the maxima in Tables 2, 3, 4, 5, are located for each subject preferentially on a few combination methods. Thus one can hope that at least for a given subject a power extraction technique, a combination method and a classifier can be determined preferentially. This work has to be done off-line, and the selected method for that subject can then be applied online. Future work would be necessary in order to define rules and methods for extracting automatically the best combination of features.

## 6 Conclusion

This study has introduced two new features for Brain-Computer Interface design: multifractal cumulants and predictive complexity. The information contained in the multifractal cumulants feature corresponds to a relation between frequency bands, rather than the power extracted in each band. The complexity feature measures the difficulty to predict the future of the EEG signals based on their past.

This study has also reproduced and extended the comparisons performed in [15] to identify the best method for extracting the power in given frequency bands, for BCI purposes. We have evaluated the two novel features and the features obtained with the various band power estimation methods on motor imagery data from several EEG data sets, using a variety of classifiers exploiting different data spaces and commonly found in the BCI literature [19]. Our results contrast with [15] in that we showed there is no best-for-all winner technique for extracting the power information. This dependence on the subject is a known characteristic in the BCI research field. Similarly, we highlighted the difficulty to use cross-validation as an indicator for the performance on the test set.

Interestingly enough, our results showed that the two new features, i.e., multifractal cumulants and predictive complexity measure, could indeed be used by themselves to discriminate between different motor imagery mental states as measured by EEG. Moreover, our results showed that when combining these two features together or with band power features, the resulting BCI could reach a better classification accuracy than any of these features alone. Thus, this suggests that these new features are a good complement to currently used features for BCI design and that they can lead to improved BCI design. Therefore we would recommend BCI designers to consider these two features as additional features in the conception of a BCI system, in order to obtain better performance.

Future work might involve the exploration of novel ways to combine the features and feature selection. Work is also needed on the design of novel algorithms including physiologically relevant error functions for EEG signal predictions for the complexity feature.

## Appendix: Source code, data, and web information

All the results in this study are reproducible independently. The code for the experiments is provided as Free/Libre software on the main author web site:

- <http://nicolas.brodu.numerimoire.net/en/recherche/publications/>

The data that was used can be downloaded on the BCI competitions website and on the OpenViBE project page:

- <http://www.bbc.de/competition/>
- <http://openvibe.inria.fr/?q=datasets>

## References

- [1] Schlögl A., Neuper C., and Pfurtscheller G. Estimating the mutual information of an eeg-based brain-computer-interface. *Biomedizinische Technik*, 47(1-2):3–8, 2002.
- [2] Patrice Abry, Stéphane Jaffard, and Bruno Lashermes. Revisiting scaling, multifractal, and multiplicative cascades with the wavelet leader lens. In *SPIE*, pages 103–117, 2004.
- [3] David P. Anderson. Boinc: A system for public-resource computing and storage. In *5th IEEE/ACM International Workshop on Grid Computing.*, November 2004.
- [4] François Auger, Patrick Flandrin, Paulo Gonçalves, and Olivier Lemoine. The time-frequency toolbox tutorial. Technical report, CNRS (France), Rice University (USA), 1995-1996.
- [5] A. Bashashati, M. Fatourehchi, R. K. Ward, and G. E. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R35–57, 2007.
- [6] B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Biomedical Engineering*, 51(6):1044–1051, 2004.

- [7] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. D. R. Millan, M. Schroder, and N. Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006.
- [8] Nicolas Brodu. Quantifying the effect of learning on recurrent spiking neurons. In *IJCNN*, pages 512–517, 2007.
- [9] Nicolas Brodu. Multifractal feature vectors for brain-computer interfaces. In *IJCNN*, pages 2883–2890, 2008.
- [10] Nicolas Brodu. Decisional states. <http://arxiv.org/abs/0902.0600>, submitted, 2009.
- [11] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [12] James Crutchfield and Karl Young. Inferring statistical complexity. *Physical Review Letters*, 62(2):105–108, 1989.
- [13] G. Dornhege, B. Blankertz, G. Curio, and K.R. Müller. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Transactions on Biomedical Engineering*, 51(6):993–1002, 2004.
- [14] Patrick Flandrin. *Temps-Fréquence*. Hermès, Paris, 1998.
- [15] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4), August 2008.
- [16] B. Hjorth. An on-line transformation of eeg scalp potentials into orthogonal source derivations. *Electroencephalography and Clinical Neurophysiology*, 39:526–530, 1975.
- [17] Kristina Klinkner, Cosma Rohilla Shalizi, and Marcelo Camperi. Measuring shared information and coordinated activity in neuronal networks. *Advances in Neural Information Processing Systems*, 18:667–674, 2006.
- [18] S. Lemm, C. Schafer, and G. Curio. BCI competition 2003–data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.
- [19] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13, 2007.
- [20] Benoit Mandelbrot, Adlai Fisher, and Laurent Calvet. A multifractal model of asset returns. Technical Report 1164, Cowles Foundation, 1997.
- [21] D. J. McFarland, C. W. Anderson, K.-R. Muller, A. Schlogl, and D. J. Krusienski. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006.
- [22] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalographic Clinical Neurophysiology*, 103(3):386–394, 1997.
- [23] Jean-François Muzy, Emmanuel Bacry, and Alain Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E.*, 47:875–884, 1993.
- [24] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *proceedings of the IEEE*, 89(7):1123–1134, 2001.
- [25] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. Openvibe: An open-source software platform to design, test and use brain-computer interfaces in real and virtual environments. *Presence: teleoperators and virtual environments*, 19(1), 2010.

- [26] Cosma Rohilla Shalizi. *Causal Architecture, Complexity, and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Wisconsin at Madison, 2001.
- [27] K. Steiglitz, T. W. Parks, and J. F. Kaiser. Meteor: A constraint-based fir filter design program. *IEEE Transactions on Signal Processing*, 40(8):1901–1909, 1992.
- [28] Herwig Wendt, Patrice Abry, and Stéphane Jaffard. Bootstrap for empirical multifractal analysis. *IEEE Signal Processing Magazine*, (38), 2007.
- [29] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.
- [30] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, 2004.