10 years of collaborations with ITT: Some contributions in evolutionary computation and applications

**Pierrick Legrand** 

Inria, CQFD Team IMB, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 Université de Bordeaux

December 9, 2019



Ínnía -



#### Management of research projects

		<b>v · · ·</b>	ANR – RNTS HEVEA ANR BNPSI
	PSI - CNRS	Psychology and sound interactions	ARC M2A3PC Micro-Projet GIS Albatros x 2
2008-2011		Fréderique Faita, Joseph Larralde, Pierre-Henry Vulliard, Myriam Desainte-Catherine. Internship Mathieu Carpentier	
	PSI – REGION	Reduction of dimension in supervised learning. Application to the study of brain activity	
	2010-2014	PhD Laurent Vezard (with Marie Chavent, Frederique Faita) Internships : Julien Clauzel, Nidal El Yacoubi and Emilie Drouineau	
	European Project	Analysis and Classification of mental states of vigilance with evolutionary computation	4 countries, 50 months of mobility 31 journal papers, 28 proceedings, 2 books 1 book chapter
	MARIE CURIE 2013-2016	PhD Yuliana Martinez, PhD Enrique Naredo, PhD Emigdio Flores, Internships : Victor Lopez Lopez, Uriel Lopez Islas, Enrique Hernandez and Luis Herrera	7 PhD defenses
	HUMO Micro-projects	Human monitoring (x 3)	
	GIS ALBATROS 2015-2017	Jean-Marc André, Eric Grivel, Frederique Faïta, Veronique Lespinet, Liliana Audin-Garcia. Internship Vincent Lenhardt, internship Luis Herrera. Starting point for the CIFRE PhD of Bastien Berthelot	
	Micro-Doppler Micro-project	Apport de l'analyse temps-frequence pour l'estimation de de micro-doppler	
(	GIS ALBATROS 2017	Eric Grivel. Internship Sabrina Macchour Starting point of the CIFRE PhD of Jimmy Bondu	
	Fraci ab toolbox	Matlab toolbox for multifractal analysis and signal processing	
	1999- ?	https://project.inria.fr/fraclab/	

Participation

#### Management of research projects

		- · · ·	ANR – RNTS HEVEA ANR BNPSI
		Psychology and sound interactions	ARC M2A3PC Micro-Projet GIS Albatros x 2
2008-2011		Fréderique Faita, Joseph Larralde, Pierre-Henry Vulliard, Myriam Desainte-Catherine. Internship Mathieu Carpentier	
	PSI – REGION	Reduction of dimension in supervised learning. Application to the study of brain activity	
	2010-2014	PhD Laurent Vezard (with Marie Chavent, Frederique Faita) Internships : Julien Clauzel, Nidal El Yacoubi and Emilie Drouineau	
	European Project	Analysis and Classification of mental states of vigilance with evolutionary computation	4 countries, 50 months of mobility 31 journal papers, 28 proceedings, 2 books 1 book chapter
	MARIE CURIE 2013-2016	PhD Yuliana Martinez, PhD Enrique Naredo, PhD Emigdio Flores, Internships : Victor Lopez Lopez, Uriel Lopez Islas, Enrique Hernandez and Luis Herrera	7 PhD defenses
	HUMO Micro-projects	Human monitoring (x 3)	
	GIS ALBATROS 2015-2017	Jean-Marc André, Eric Grivel, Frederique Faïta, Veronique Lespinet, Liliana Audin-Garcia. Internship Vincent Lenhardt, internship Luis Herrera. Starting point for the CIFRE PhD of Bastien Berthelot	
	Micro-Doppler Micro-project	Apport de l'analyse temps-frequence pour l'estimation de de micro-doppler	
1	GIS ALBATROS 2017	Eric Grivel. Internship Sabrina Macchour Starting point of the CIFRE PhD of Jimmy Bondu	
	Fraci ab toolbox	Matlab toolbox for multifractal analysis and signal processing	FreeLeb <sub>20</sub>
	1999- ?	https://project.inria.fr/fraclab/	

Participation









Theory



















## PART 1: Artificial Evolution

- Definitions
- Prediction of expected performance for GP

## PART 2: Estimation of signal regularity

Hölderian regularity

## PART 3: Applications

- Evolutionary computation for EEG classification
- Regularity estimation with Genetic Programming

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming

## PART 1: Artificial Evolution

Inria, CQFD Team IMB, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 Université de Bordeaux

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming

A set of techniques grouped under a generic term Ingredients Evolutionary loop Example

### **Genetic Algorithms**

Discrete representation: Genetic Algorithms

## **Evolution strategies**

Continuous representation: Evolution Strategies

## **Genetic Programming**

Functional representation: Genetic programming Example: Using GP for regression

Artificial Darwinism ○●○○○○	Genetic Algorithms	Evolution Strategies	Genetic Programming

Stochastic optimization which uses mechanisms inspired by the biological evolution, such as:

- reproduction,
- mutation,
- selection and
- survival of the strongest individuals

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming		
00000					
A set of techniques grouped und	A set of techniques grouped under a generic term				

Evolutionary Algorithms	Genetic Algorithms (GA)
	Evolution Strategies (ES)
	Genetic Programming (GP)

Artif	icial	Darwi	nism

Ingredients

#### Genetic Algorithms

Evolution Strategies

Genetic Programming



**Genetic Operators** 

Pierrick Legrand IMB/INRIA/UBX

Artificial Darwinism ○○○○●○	Genetic Algorithms	Evolution Strategies	Genetic Programming
Evolutionary loop			



Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming
Example			

Potential Solutions = Individuals in a population



Artificial Darwinism	Genetic Algorithms ●○	Evolution Strategies	Genetic Programming

A set of techniques grouped under a generic term Ingredients Evolutionary loop Example

## **Genetic Algorithms**

Discrete representation: Genetic Algorithms

## **Evolution strategies**

Continuous representation: Evolution Strategies

## **Genetic Programming**

Functional representation: Genetic programming Example: Using GP for regression

Artificial Darwinism	Genetic Algorithms ○●	Evolution Strategies	Genetic Programming
Discrete representation: Genetic	Algorithms		

Each individual is represented by a binary string.

John H. Holland (1960, 1975), David Goldberg (1989)





Mutation

Artificial Darwinism	Genetic Algorithms	Evolution Strategies ●○	Genetic Programming

A set of techniques grouped under a generic term Ingredients Evolutionary loop Example

### Genetic Algorithms

Discrete representation: Genetic Algorithms

## **Evolution strategies**

Continuous representation: Evolution Strategies

## **Genetic Programming**

Functional representation: Genetic programming Example: Using GP for regression

Artificial Darwinism	Genetic Algorithms	Evolution Strategies ○●	Genetic Programming
Continuous representation: Evo	lution Strategies		

Each individual is a vector in  $\mathbb{R}^n$ .

Hans-Paul Schwefel (1970)

#### Barycentric crossover

 $\begin{aligned} \forall i \in \{1,..,n\}, x_i^{children} &= \alpha x_i^{father} + (1-\alpha) x_i^{mother} \\ \alpha \text{ random value in } [-\epsilon,1+\epsilon]. \end{aligned}$ 

#### Gaussian mutation

 $\forall i \in \{1,..,n\}, x_i^{children} = x_i^{children} + N(0,\sigma)$  Two parameters  $P_m$  and  $\sigma$ .

Artificial Darwinism	Genetic Algorithms	Evolution Strategies ○●	Genetic Programming
Continuous representation: Evo	lution Strategies		

Each individual is a vector in  $\mathbb{R}^n$ .

Hans-Paul Schwefel (1970)

#### Barycentric crossover

 $\begin{aligned} \forall i \in \{1,..,n\}, x_i^{children} &= \alpha x_i^{father} + (1-\alpha) x_i^{mother} \\ \alpha \text{ random value in } [-\epsilon,1+\epsilon]. \end{aligned}$ 

#### Gaussian mutation

 $\forall i \in \{1,..,n\}, x_i^{children} = x_i^{children} + N(0,\sigma)$  Two parameters  $P_m$  and  $\sigma$ .

#### Demo

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming ●○○

A set of techniques grouped under a generic term Ingredients Evolutionary loop Example

## **Genetic Algorithms**

Discrete representation: Genetic Algorithms

### **Evolution strategies**

Continuous representation: Evolution Strategies

## **Genetic Programming**

Functional representation: Genetic programming Example: Using GP for regression

Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming ○●○
Functional representation: Gene	tic programming		

#### Definition

Genetic programming (GP) is an evolutionary computation (EC) technique that automatically solves problems without requiring the user to know or specify the form or structure of the solution in advance [Koza, 1992].



Artificial Darwinism	Genetic Algorithms	Evolution Strategies	Genetic Programming ○○●
Example: Using GP for regression	on		

#### Symbolic Regression

Given a set of input data X and a set of desired outputs Y, find a function f such that:  $f(X_i) = Y_i \quad \forall i \in \{1, N\}$ 



Introduction	PEP	SPEP	Conclusion

## Prediction of expected performance for GP

This work, related to the PhD thesis of Yuliana Martínez (ITT Tijuana) and developed in the context of the European ACOBSEC project, has been published in Genetic Programming and Evolvable Machine, Springer Verlag, 2016, 17 (4), pp.409-449. Work carried out with Yuliana Martínez, Leonardo Trujillo and Edgar Galván-López.

Introduction	PEP	SPEP	Conclusion

Introduction	PEP	SPEP	Conclusion
0000			

## Introduction

## Prediction of Expected Performance (PEP)

Proposal Feature Extractio

Synthetic classification problems and classification error

Correlation between features and classification error

Building PEP model

Testing PEP models in synthetic classification problems

Testing PEP models in real-world classification problems

## Specialist Prediction of Expected Performance (SPEP)

Proposal SPEP using two groups

## Conclusion

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
○●○○	000000000000	000	

#### **Expected Performance Prediction**

- Research in Evolutionary Computation (EC) has produced many flexible and robust problem solving algorithms.
- However, in many areas, particularly Genetic Programming (GP), it's not yet clear if a particular algorithm will perform well on an specific problem.
- Therefore, it would be desirable to be able to grade each problem based on its difficulty.
- Such a grade will depend upon the solution method used. In this case we will use Genetic Programming (GP).

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
○○●○	000000000000	000	

In GP search, several works have attempted to determine the difficulty that a problem poses. Two broad groups of methods are available.

Introduction	<b>PEP</b> 000000000000	<b>SPEP</b> 000	Conclusion

In GP search, several works have attempted to determine the difficulty that a problem poses. Two broad groups of methods are available.

**Evolvability Indicators** (EI), focuse their analysis on the fitness landscape and how it relates to the difficulty of a search [Altenberg, 1994, Vanneschi et al., 2007, Poli and Vanneschi, 2007, Tomassini et al., 2005, O'Neill et al., 2010, McDermott et al., 2010, Malan and Engelbrecht, 2013].

In GP search, several works have attempted to determine the difficulty that a problem poses. Two broad groups of methods are available.

- Evolvability Indicators (EI), focuse their analysis on the fitness landscape and how it relates to the difficulty of a search [Altenberg, 1994, Vanneschi et al., 2007, Poli and Vanneschi, 2007, Tomassini et al., 2005, O'Neill et al., 2010, McDermott et al., 2010, Malan and Engelbrecht, 2013].
- 2 Predictors of Expected Performance (PEP), characterize problem difficulty using the problem domain as the frame of reference and to measure problem difficulty based on the expected performance of the GP search, derived using a domain specific description of each problem.

[Graff and Poli, 2010, Graff and Poli, 2011, Graff et al., 2013, Trujillo et al., 2011a, Trujillo et al., 2011b, Trujillo et al., 2011c].

Introduction	PEP	SPEP	Conclusion
0000			

#### Why not use Evolvability Indicators?

# Fitness landscape, epistasis, neutrality, locality, Fitness Distance Correlation (FDC), Negative Slope Coefficient (NSC), fitness cloud

- It is necessary to execute the evolutionary process.
- In GP unlike GA, to represent the fitness landscape is a difficult task.
- A comparative study between EI and PEP, presented in [Martinez et al., 2012], showed that GP-PEP models are more correlated with the classification error than the NSC measure.
| Introduction | PEP         | SPEP | Conclusion |
|--------------|-------------|------|------------|
|              | 00000000000 |      |            |

# Introduction

## Prediction of Expected Performance (PEP)

Proposal Feature Extraction Synthetic classification problems and classification error Correlation between features and classification error Building PEP model Testing PEP models in synthetic classification problems Testing PEP models in real-world classification problems

### Specialist Prediction of Expected Performance (SPEP) Proposal SPEP using two groups

# Conclusion

Introduction	<b>PEP</b> ○●○○○○○○○○○○	<b>SPEP</b> 000	Conclusion
Proposal			





Introduction	<b>PEP</b> ○○● <b>○○○</b> ○○○○○○○	<b>SPEP</b> 000	Conclusion
Feature Extraction			

### Geometric mean (SD):

Measures the homogeneity of covariances [Michie1994,So1999].

$$SD = exp\left\{\frac{M}{m\sum_{i=1}^{C}(n_i - 1)}\right\}$$

where *C* is the number of classes,  $n_i$  is the number of the instances for *i*-th class and *m* is the number of attributes in the data.  $S_i$  and *S* are the unbiased estimators of the  $i^{th}$  sample covariance matrix and the pooled covariance matrix respectively.  $n = \sum_{i=1}^{C} n_i$ .

$$M = \gamma \sum_{i=1}^{C} (n_i - 1) \log |S_i^{-1}S|$$

$$\gamma = 1 - \frac{2m^2 + 3m - 1}{6(m+1)(C-1)} \sum_{i=1}^{C} \left(\frac{1}{n_i - 1} - \frac{1}{n - C}\right)$$

$$S = \frac{1}{n - C} \sum_{i=1}^{C} (n_i - 1) S_i$$



Feature Efficiency (FE)

### Feature Efficiency (FE):

Measures the amount by which each dimension contributes to the separation of both classes. This measure is computed for the  $j^{th}$  dimension by

$$FE_j = \left(1 - \frac{\eta_j}{n}\right)$$

where  $\eta_j$  represent the number of points inside the overlapping region and n is the total number of sample points. Finally,  $FE = max(\{FE_j\})$  with j integer in [1, m].

Introduction	<b>PEP</b> ○○ <b>○○●○○</b> ○○○○○○	<b>SPEP</b> 000	Conclusion
Feature Extraction			



### Class Distance Ratio (CDR):

Compares the dispersion within the classes to the gap between the classes [Ho2002]. For each data sample, compute the Euclidean distance to its nearest neighbor within the class (intraclass distance) and nearest-neighbor from the other class (interclass distance). The CDR is the ratio of the averages of all intraclass and interclass distances.



Volume of Overlap Region (VOR)

X1

### Volume of Overlap Region (VOR):

Provides an estimate of the amount of overlap between both classes [Ho2002]. The VOR is computed by finding, for each dimension, the maximum and minimum value of each class and then calculating the length of the overlap region. The length obtained from each dimension is then multiplied to measure the overlapping region. The VOR is zero when there is at least one dimension in which the two classes do not overlap.

Introduction	<b>PEP</b> ○○ <b>○○○</b> ●○○○○○○	<b>SPEP</b> 000	Conclusion
Feature Extraction			

### Canberra Distance (CD):

Provides a numerical measure of the distance between pairs of points in a vector space. Suppose a problem has m dimensions, we take a rank statistic of the samples of each class, call it  $x_i$  for class 1 and  $y_i$  for class 2, for the *i*-th dimension. This produces two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , such that  $\mathbf{x} = (x_1, ..., x_m)$  and  $\mathbf{y} = (y_1, ..., y_m)$ . The CD is given by:

$$CD(\mathbf{x}, \mathbf{y}) = \frac{1}{m} \sum_{i=1}^{m} \frac{|x_i - y_i|}{|x_i| + |y_i|}.$$

In this work, we use the CD to describe the distance between both classes using three rank statistics: (1) CD-1 uses the 1st quartile; (2) CD-2 uses the median; and (3) CD-3 uses the 3rd quartile.

Introduction	PEP	SPEP	Conclusion
	000000000000		
Synthetic problems			



Examples of synthetic classification problems, specifying the CE $\mu$  and standard deviation  $\sigma$  achieved by PGPC. These ordered from the lowest CE $\mu$  (easiest) to the highest CE $\mu$  (hardest).





Relationship between the CE $\mu$  (x-axis) and each descriptive feature (y-axis) for all problems  $p \in Q$ , where  $\rho$  specifies Pearson's correlation coefficient.

Introduction	PEP ○○○○○○○○● <b>○</b> ○○	<b>SPEP</b> 000	Conclusion
Building PEP model			

Three versions of the problem are posed, each with a different terminal set defined as subsets of all extracted features(4F, 5F, 7F).

- Set 4F uses the features with the four highest correlation coefficients (FE, CDR, VOR and CD-1),
- set 5F uses the features with the five highest correlation coefficients (SD, FE, CDR, VOR and CD-1),
- and 7F uses all of the seven features.

Introduction	<b>PEP</b> ○○○○○○○○ <b>○</b> ●○○	<b>SPEP</b> 000	Conclusion
Building PEP model			







(d) [PEP-4F:  $\rho = 0.9634$ ] (e) [PEP-5F:  $\rho = 0.9688$ ] (f) [PEP-7F:  $\rho = 0.9636$ ]

Performance prediction of the best PEP models evolved with the different feature set: PEP-4F(left), PEP-5F(middle) and PEP-7F(right). First line: PCE of the best solution and the know CE $\mu$ . Second line: scatter plots of the PCE and the CE $\mu$ .



 $(d) [PEP-4F: \rho = 0.8634] (e) [PEP-5F: \rho = 0.8823] (f) [PEP-7F: \rho = 0.8046]$ 

Performance prediction of the best PEP models evolved with the different feature set. PEP-4F (left), PEP-5F (middle) and PEP-7F (right). First line: PCE of the best solution and the know CE $\mu$ . Second line: scatter plots of the PCE and the CE $\mu$ .

Introduction	PEP	SPEP	Conclusion
		000	

# Introduction

## Prediction of Expected Performance (PEP)

Proposal Feature Extraction Synthetic classification problems and classification error Correlation between features and classification error Building PEP model Testing PEP models in synthetic classification problems Testing PEP models in real-world classification problems

### Specialist Prediction of Expected Performance (SPEP) Proposal SPEP using two groups

# Conclusion

Introduction	<b>PEP</b> 000000000000	SPEP ○●○	Conclusion
Proposal			



Introd	uction
0000	

SPEP

#### SPEP using two groups



Performance prediction of the best Ensemble-2 solutions for each feature set: 4F (left), 5F (middle) and 7F (right). First line: ground truth  $CE_{\mu}$  of each problem (triangles) and the corresponding PCE (circles). Second line: scatter plots between the  $CE_{\mu}$  and the corresponding PCE. The PCE is presented in three different cases: (1) the PCE of a correctly classified problem (CC-PCE, circle); (2) the PCE of a misclassified problem (MC-PCE, dark circle); and (3) the oracle PCE of a misclassified problem using the correct SPEP (O-PCE, circle with a cross).

Introduction	PEP	SPEP	Conclusion
			00

# Introduction

### Prediction of Expected Performance (PEP)

Proposal Feature Extraction Synthetic classification problems and classification error Correlation between features and classification error Building PEP model Testing PEP models in synthetic classification problems Testing PEP models in real-world classification problems

# Specialist Prediction of Expected Performance (SPEP)

Proposal SPEP using two groups

# Conclusion

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
	000000000000	000	○●

• The proposed models predict the performance of the GP classifier when they are evaluated on the test set of fitness cases.

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
	0000000000000	000	○●

- The proposed models predict the performance of the GP classifier when they are evaluated on the test set of fitness cases.
- An ensemble of SPEPS built for each group improving the prediction accuracy.

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
	000000000000	000	○●

- The proposed models predict the performance of the GP classifier when they are evaluated on the test set of fitness cases.
- An ensemble of SPEPS built for each group improving the prediction accuracy.
- This methodology can be used for many classifiers and then build an expert system for classifier selection.

Introduction	<b>PEP</b>	<b>SPEP</b>	Conclusion
	000000000000	000	○●

- The proposed models predict the performance of the GP classifier when they are evaluated on the test set of fitness cases.
- An ensemble of SPEPS built for each group improving the prediction accuracy.
- This methodology can be used for many classifiers and then build an expert system for classifier selection.
- This methodology could be extended to GP-based symbolic regression.

# PART 2: Estimation of signal regularity



# Hölderian regularity

HDR Legrand IMB/INRIA/UBX

# Definitions

Signal Regularity

# Estimation

Oscillations Regression of wavelet coefficients

### Application Estimation on synthetic signal

### Definitions Signal Regularity

# Estimation

Oscillations Regression of wavelet coefficients

Application Estimation on synthetic signa



Definitions ○●	Estimation	Application
Signal Regularity		
The Hölder pointwise exponent	is the most common tool used to	measure the

The Hölder pointwise exponent is the most common tool used to measure the regularity of a signal at a given point.

#### Definition

Let f be a function from  $\mathbb{I}$  to  $\mathbb{I}$ , s > 0,  $s \in \mathbb{I} \setminus \mathbb{N}$  and  $x_0 \in \mathbb{I}$ . Then  $f \in C^s(x_0)$  if and only if there is a real  $\eta > 0$ , a polynomial P of degree smaller than s and a constant c such that

 $\forall x \in B(x_0, \eta), \quad |f(x) - P(x - x_0)| \le c|x - x_0|^s$ 

By definition, the pointwise exponent of f at  $x_0$ , noted  $\alpha_p(x_0)$  is the supremum of s such as  $f \in C^s(x_0)$ .



Hölderian envelope of a signal at the point  $x_0$ .

Definitions Signal Regularity

### Estimation Oscillations

Regression of wavelet coefficients

Application Estimation on synthetic sign



Definitions	Estimation ○●○○○	Application
Oscillations		

A function f(t) is Hölderian of exponent  $\alpha \in [0.1[$  at t if there is a constant c such that for any t' in a neighbourhood of t,

$$|f(t) - f(t')| \le c|t - t'|^{\alpha}$$

In terms of oscillations, this condition can be written:

A function f(t) is Hölderian of exponent  $\alpha$  at t, with  $0 < \alpha < 1$  if there is a constant c such that for any  $\tau$ ,

$$osc_{\tau}(t) \leq c\tau^{c}$$

with

$$osc_{\tau}(t) = \sup_{|t-t'| \le \tau} f(t') - \inf_{|t-t'| \le \tau} f(t') = \sup_{t', t'' \in [t-\tau, t+\tau]} |f(t') - f(t'')|$$

Then the regularity estimator will be constructed at each point as **the slope of the** regression of the logarithm of the oscillation as a function of the size of the ball in which the oscillation is calculated.

### Theorem (S. Jaffard)

$$|c_{j,k}| \le c2^{-j(\alpha + \frac{1}{2})}(1 + |2^j t_0 - k|)^{\alpha} \quad \forall j, k \in \mathbb{Z}^2$$

Conversely ;

$$If \ \forall j,k \in \mathbb{Z}^2 \ one \ has \ |c_{j,k}| \leq \ c2^{-j(\alpha+\frac{1}{2})}(1+|2^jt_0-k|)^{\alpha'}$$

for a  $\alpha' < \alpha$  then, the Hölder exponent of f in  $t_0$  is  $\alpha$ .

Definitions	Estimation ○○○●○	Application
Regression of wavelet coefficients		



Regression calculated over a point of the signal. Left image shows a dyadic wavelet decomposition, and the right image display the actual regression calculated over the point  $t_0$ , where each dot corresponds to each  $log_2$  of the wavelet coefficient magnitude located above  $t_0$ .

Definitions oo	Estimation ○○○○●	Application
Regression of wavelet coefficients		

### Definition

At each point  $t_0$  of the signal, the regularity is estimated by:

$$\alpha(n,t_0) = -p - \frac{1}{2}$$

with p the slope of the least square linear regression of the logarithms of the wavelet coefficients "above" this point as a function of the scales.

Definitions ○○	Estimation ○○○○●	Application
Regression of wavelet coefficients		

#### Definition

At each point  $t_0$  of the signal, the regularity is estimated by:

$$\alpha(n,t_0) = -p - \frac{1}{2}$$

with p the slope of the least square linear regression of the logarithms of the wavelet coefficients "above" this point as a function of the scales.

### Theorem

At each point  $t_0$  of the signal decomposed on n scales, we estimate the regularity by the following formula:

$$\alpha(n, t_0) = -\frac{1}{2} - K_n \sum_{j=1}^n s_j \log_2 |c_{j,k}|$$

with  $K_n = \frac{12}{n(n-1)(n+1)}$  et  $s_j = j - \frac{n+1}{2}$ . The  $c_{j,k}$  are the wavelet coefficients above  $t_0$ .

We note k but the value is  $\lfloor \frac{t_0+1}{2^{n-j+1}} \rfloor.$ 

#### Definitions Signal Regular

Signal Regularity

# Estimation

Oscillations Regression of wavelet coefficients

Application Estimation on synthetic signal Estimation

Estimation on synthetic signal



# PART 3: Applications

# Evolutionary computation for EEG classification

This work is related to the PhD thesis of Laurent Vezard and developed in the context of the PSI Region Project and the ACOBSEC European project. A slightly different version has been published in a book chapter. Eduardo Miranda; Julien Castet; Benjamin Knapp. Guide to Brain-Computer Music Interfacing, Springer, 2014. Work carried out with Laurent Vézard, Marie Chavent, Frédérique Faïta-Ainseba and Leonardo Trujillo.
EEG data Acquisition Acquisition Protocole

Feature Extraction Slope Criterion

Evolutionary Algorithm Design Results

# Goal

- Characterize the state of alertness of a person from his electroencephalogram (EEG).

EEG data Acquisition Acquisition Protocole

Feature Extraction Slope Criterion

Evolutionary Algorithm Design Results

- First EEG recording: subject in a normal state of alertness: "normal"
- Second EEG recording: subject in a state of low vigilance: "relax"



Feature Extraction

Evolutionary Algorithm



- EEG headset installation time: 45 minuts.
- Subject with open eyes.
- Sampling frequency: 256Hz.
- Recording time: 3 minuts (46000 sample points).

### Campaigns:

- 58 electrodes renumbered from 1 to 58
- Subjects under 35, right-handed and non-smoker
- 58 subjects  $\Rightarrow$  16 preserved

### **Relaxation session**

20 minutes with a recorded voice offering 3 exercises:

- Autogenic training [Schultz1958]: repetition of sentences, self-hypnosis.
- Progressive muscle relaxation [Jacobson1974].
- Mental visualization (familiar places, smells, noises).

Feature Extraction

Evolutionary Algorithm

### 3 minutes of EEG recording before relaxation.

FP1-Ref	man man hand and and the	and a second production of the second second second and the second s
FPz-Bet	materialmentanten	and a second where the second and the second and the second and the second and the second second second and the second seco
FP2-Ref	man and an and a man and an	alo and an and the second the second of the second and a second and a second and the second and the second of the
AF3-Ret	man man in marine man	ward and an and the second and a second and the sec
AF4-Ret	we have have been and the second	and a particular a second a
F7-Bet	where where the second water	and all and a second and a second a second a second a second a second and a second as a second a second a second
F5Ref	and the second and the second	ad a when a second a
F3Bet	www.warnerson	and the second
FTHM	and maker and and and	
Fallet	m man contractor	
F 2198	the stand of the stand	
14198	and an est of an estimate adam	
FORM	the A marked and amounth addition	
ETTERAL	Burn Allertrans harded make when the	and the second
EL3 DA	no Manual white the the	and the second
FC1-Ref	m. Moner function	and the second second and the second
ED: Bet	m. march unhan	and a faith and a faith and a second a second and a second a
FC2-Rel	mi magal under	and the second and the
FC4Rel	man mager happender that the	and the first of the second of the second second and the second second second second second second second second
FDS-Rel	many agent the second	
T3-Rel	"SAMATING AND	ኯዿዄኯዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀዀ
D5 Rel	mount internetion	and the second and a second a second and a second a
C3-Rel	www.www.www.www.www.	and the second and the
C1 Ref	an water and a series of the s	
Caller	and any and a hard a state	
CARL	have been a start and a short of	and a stand of the second stand of the second stand
CO Red	have a comparison for stand and the state of	
T & D + 4	a he was a propriet which a sold	and a real with a second and have and the second advantage of the second by a second by a second and the second advantage of t
TP7.Ref	Anna harrow and and the second	and a second second and a second second and the second second second second second second second second second
0P58ef	mangroup and makes	and and a second a second a second and a second a
DP3Ref	Mar Margaret Marchan Marchan	and a second the second and the seco
DP1-Ref	Man margare marken Margare	and the second of the second second and the second s
CP2-Ref	monther marken	and the second and the second and the second s
DP2-Ref	Carl Maria Marian	and the second
D-4-Het	and where a second second	and a farmer of the farmer and the second
CP6Her	Mary and the second second	
TPOPper	white others a back and he add	
DS D 4	man and an and an and an and	
P3-Rel	monterentermanter	and a far and a second as a second as a second
P1-Ref	Man margaren man Maryan	and a configuration of the second and
PaRef	Man margaren margane	and the second water and the second
P2-Ref	Carbon Substant Supervision	
Parter	and the second second	
TGRA	Manager and a branch the first	
PO7-Ref	when the an an intermediate	and the second and the
P05-Rel	Man Margaren Margar	and the second and the
P03-Ref	man water our warden and	and the second and the second and the second second and the second s
PO2-Ref	and the second s	
POArtiet	The pater A second March	
POB-Rel	mannentermont	And a second and the
018af	month and an and the same	and the second second and the second
02Ret	Heren marker warden and the second	the production of the second
02Ref	montroburrowsprotes	and the construction and the second and the second and the second second and a second and a second and a second and the
EUGREDOR.	had done and have been and	
62-Rel	meren part any mound	
1		
10 >4404		

Pierrick Legrand IMB/INRIA/UBX

### 3 minutes of EEG recording after relaxation.

#### Alpha waves

FP1-Ref	Manual West Strength and an and a strength and a st		u/wh	have a provide the second of the second s
FP2-Ref	Warden and a mark the second and the second	and the man to and a many more and the		and an and an and and and and and and an
AF3-Ref	many many man his with the second	man mark of the man man who	ique.	and and the second an
AF4-Bef	Man market water water and the state of the second	man man and man and man and and and and and and and and and a	appe	manual manual and a second and a
F7-Rel	West Strates and the second strates and second	and an an an water and proceeded with a few	wmw.	have a for the stand of the sta
F5-Rei	and Manus Manus and and and and the	warmenter and a second and the	chank	and the second state of the second state and the second state of t
F3-Rol	and and and a second and a second and a second	and the second of the second of the second of the	-prime	a second and as
FIRE	the second second second second second		- m	Mun manual M
E2.Del	many many many the manufacture property and	and many of the manual of the		have a support of the
FARel	man more the month and the many many	an marken and and a manager of the	-	more manager have and a more all were
F6-Rel	many any manufacture many many	man manual man manus and the manus of the		my many many Many many and and and
F8-Rei	many many many many many many many many	and a start we have the start and the second and the	in	and a second and a
FC5-Ref	mander and an and and a second and a second and and and and and and and and and a	warmen war and a second with the second with	-	an many and a super a super and a super a supe
FC3-Ref	wanter and man and water and the second	and the manufacture and the second second	11 mar	and the second and the second and the second and the
FC1-Ref	and a superior and a superior and and a superior	man man and a man a man a man	1 m	and the second and a second of the second se
FCz-Ref	the second	and the second of the second of the second of the	A.	and the second of the second o
FU2-Het	An and the second second and the second second	WWW. WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW	1000	A way a w
FL4HB	a start for the second se		A	
TODA	a manufacture demand our all all the state of the state o	Trademan de Materia Marine in alla and an and an	Long	an assure that the second as a second way have not a second and a second and the second as a second se
CERH	Man har war and the second s	a manual of the form of the second of the	and and	when the second s
C3Rel	man and the second of the seco	anon march to the a company and	1	and the second and th
C1-Bei	him have more and the second	among and a second and and	ser.	man man man man man man man man
Cz-Rel	that her minute many many many	and a man and a man and a man	5-	and a man a
C2-Rel	the man the man the second and the second	and the the way when the	ser.	a margare and the source and the
C4-Rei	much man and a second and a second	and the second water and the second s	w w	a me was a superior and a superior and
C6-Rel	When the stand of	and a second and a second and a second and a second a se	A	and and an appropriate the second
T4-Rei	have a start of the start of th	and the second	Mr.	and the support of the state of
TP7-Bet	Monthly March 100 march 100 miles and all all and all all all all all all all all all al	and the second of the second of the second	a from	per and a second provident and a second provident and the second s
CPOHM	Marshing and the second second and the second		-	the second watches and a share of the strength
CD1.D+i	the second		-	a walk a walk was a shown and and
CP+Ref	and the man more man and the second		S	- man man man man
CP2-Bet	man hopen mound menning remaining		in	man man man man man
CP4-Ref	my how more many and	and the second way that the second	w w	a a marine and a marine and a second
CP6-Ref	and have the manual and	and a marine a compared to the	st.	a man and a second
TP8-Ref	may have a man and the state and and the	and manufacture and the second	and and	a contration and a second and the se
T5-Rel	And a start water and the second of the seco		and a	a har al a second and the second and the second s
PSRe	and the second and the second of the second			and the second
D1.DV	man B. And the the the state of		S.	m and an way show a show me make
P2Rel	and the how more and the services		in	a manuna manuna manuna
P2-Rel	would be have never when and the approximation		w.	a realized and and a second and a second and and a second a
P4-Rel	and the manuscription and the second		Sur.	C. Marken Marken Marken and Marken
P6-Re	- A Maria Maria Maria Maria Maria		s.	A have a server a server a server and the server and the server a server a server a server a server a server a
1688 007.0-4		A A A A A A A A A A A A A	the second	the set of the first standard and should also all also all the set of the
POSPei	a man acture minuner much much man and	A A MAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	in	m an a commence and an and and
P03-Ref	most the for more many and	NATATA ANALANA ANALANA ANALANA ANALANA	Mr.	M M Many and Many Marine
P0z-Ref	man Man more man more many	MATATALY AVALANAVAVAVAVAVAVAVAVAVAVA	M.A	white the second white the second white the second
P04 Ref	man and and and the second		And A	Margan and and and and and and and and and a
PUSFlef	a share and a share a			
nuo-riet	and Andrew and the second	A A A A A A A A A A A A A A A A A A A	Ser.	
0z-Bei	man Arman manunan and and	and the second of the second o	sin,	and the second and the second and the second
02-Ref	man ham an an an and and and and and and and an		4m	May have marked and the second
EUGREUGL				
A1-Bel	and the second	and a second		and a second
AZHE	and the second se	Manual Manual Contraction of the second s		21
10:49:23	Sa A			
	The second			

Pierrick Legrand

IMB/INRIA/UBX

EEG data Acquisition Acquisition Protocole

### Feature Extraction Slope Criterion

Evolutionary Algorithm Design Results

EEG data Acquisition	Feature Extraction ○●○○○○	Evolutionary Algorithm
Slope Criterion		

The dyadic grid gives a spatio-frequential representation of the discrete dyadic wavelet decomposition



### - Alpha: 8 - 12Hz.

- Waves characteristics of a relaxed state.



Linear regression between 4 and 16Hz.

- Alpha: 8 12Hz.
- Waves characteristics of a relaxed state.



Linear regression between 4 and 16Hz.



Linear regression between 4 and 16Hz.

1/16

1/4

1/64

0

Frequency (Hz)

16

4

64 128

### Slope Criterion



Slope criterion, sum on subjects for each electrode

Pierrick Legrand IMB/INRIA/UBX

### **Slope Criterion**



Slope criterion, sum on electrodes for each subjects

- ⇒ Very strong inter-subject variability
- ⇒ This criterion that does not allow to build a powerful classifier for different subjects.

EEG data Acquisition

Feature Extraction ○○○○○● Evolutionary Algorithm

**Slope Criterion** 

Classification: Usual methods

	K nearest	Binary	Random	Discriminant	Sparse Discriminant
	neighbors	decision trees	forests	PLS	PLS
Mean	37.28	33.98	32.03	40.63	36.25
Standard Deviation	10.47	5.15	6.46	8.55	7.96

Mean and standard deviations of Correct Classification Rates for different classification methods applied on slope criterion.

### This approach is not efficient

**Our contribution**: Design a relevant **evolutionary algorithm** to solve this task of classification.

 $\Rightarrow$  Find the relevant electrodes.

 $\Rightarrow$  Find the relevant frequencies for the calculation of the slope criterion.

EEG data Acquisition Acquisition Protocole

Feature Extraction Slope Criterion

Evolutionary Algorithm Design Results EEG data Acquisition

Design

Feature Extraction

Evolutionary Algorithm

### Example of a genome in the evolutionary algorithm



EEG	data	Acquisition

Feature Extraction

Evolutionary Algorithm

#### Design

### Relationship between the genome and the calculation of the slope criterion



EEG data	Acquisition

Feature Extraction

Results

### Average correct classification rate

Evaluation	CCR		
Method	Mean	Standard deviation	
CART	86.68	1.87	
SVC	83.49	2.37	

Average and standard deviations of the correct classification rates obtained for the 100 runs of the evolutionary algorithm and for two methods of evaluation.

#### Results

### Best genome

Evaluation	BEST genome			
method	Selected	Selected	Correct classification	
	electrode	frequency (Hz)	rates	
CART	F4	1/8, 1/4, 2, 4 et 64	89,33%	
SVC	F2	1/32, 1/16, 2, 4, 8, 64 et 128	89,33%	

Table summarizing the two best genomes found during the 100 runs of the genetic algorithm with two methods of evaluation.

# Regularity estimation with Genetic Programming

Joint work with Leonardo Trujillo, Gustavo Olague and Jacques Levy-Vehel. Evolving estimators of the pointwise Hölder exponent with Genetic Programming. Information Sciences 209 (Nov. 2012), 61-79. Hölderian Regularity

Contribution Training set

Results

Hölderian Regularity

Contribution Training set

Results

Pierrick Legrand IMB/INRIA/UBX

Contribution	

Hölder exponent

Mathematical tool that measures the regularity of a signal around each point.





Hölderian Regularity	Contribution	Results
000		

### General motivation



$$\alpha_p(x_0) = \liminf_{h \to 0} \frac{\log |f(x_0 + h) - f(x_0)|}{\log |h|}$$

Hölderian envelope of signal f at point  $x_0$ 

• For real-world signals the Hölder exponent must be estimated for each point.

Hölderian Regularity	Contribution	Results
000		

### General motivation





Hölderian envelope of signal f at point  $x_0$ 

- For real-world signals the Hölder exponent must be estimated for each point.
- Several estimation methods exist, but most methods are slow or highly parameterized;

Hölderian Regularity	Contribution	Results

### General motivation



$$\alpha_p(x_0) = \liminf_{h \to 0} \frac{\log |f(x_0 + h) - f(x_0)|}{\log |h|}$$

Hölderian envelope of signal f at point  $x_0$ 

- · For real-world signals the Hölder exponent must be estimated for each point.
- Several estimation methods exist, but most methods are slow or highly parameterized;
- Therefore there use is not common (particularly in applications where speed can be of importance)

Hölderian Regularity

# Contribution Training set

Results

• Evolve estimators of the pointwise Hölder exponent for 2D signals with Genetic Programming.

- Evolve estimators of the pointwise Hölder exponent for 2D signals with Genetic Programming.
- GP evolves estimators that are **accurate** and **fast**.

- Evolve estimators of the pointwise Hölder exponent for 2D signals with Genetic Programming.
- GP evolves estimators that are **accurate** and **fast**.
- Evolution is a one-shot process, evolved estimators can be used easily.

- Evolve estimators of the pointwise Hölder exponent for 2D signals with Genetic Programming.
- GP evolves estimators that are **accurate** and **fast**.
- · Evolution is a one-shot process, evolved estimators can be used easily.

### Awards

- Best Paper Award in the track Genetic Programming, GECCO 2010, Portland, Oregon.
- Humies Award Finalist, GECCO 2013, Amsterdam, The Netherland.



We generate three groups of images with **FracLab**, using three different functions that take as input the point coordinates (x, y) of an image and provide as output the desired regularity; these functions are:

1 a Polynomial 
$$p_1(x, y) = 0.1 + 0.8xy;$$

**2** a Sine 
$$p_2(x,y) = 0.5 + 0.2(sin(2\pi x))(cos(\frac{3}{2}\pi y));$$

3 an *Exponential* 
$$p_3(x,y) = 0.3 + \frac{0.3}{1+e^{-100(x-0.7)}}$$
.

These functions provide the prescribed regularity needed to build the synthetic images used for training and testing of our evolved operators.

Training set

(a) Polynomial  $p_1$ 





(c) Exponential  $p_3$ 

Training set

#### Contribution ○○○○●



(d) Polynomial

(e) Sine

(f) Exponential

These images have a prescribed regularity given by functions  $p_1$  (Polynomial),  $p_2$  (Sine) and  $p_3$  (Exponential).
Hölderian Regularity

Contribution Training set

Results

Contribution

Results ○●

## **Results: Real Images**



**Original Image** 



**GP-Estimator** 



**Traditional Method** 



**GP-Estimator**