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

Pierrick Legrand

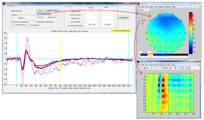
Inria, CQFD Team
IMB, Institut de Mathématiques de Bordeaux, UMR CNRS 5251
Université de Bordeaux

December 9, 2019

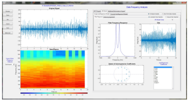
Management of research projects

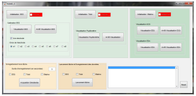
PSI – CNRS 2008-2011	**Psychology and sound interactions** Fréderique Faïta, Joseph Larralde, Pierre-Henry Vulliard, Myriam Desainte-Catherine. **Internship Mathieu Carpentier**
PSI – REGION 2010-2014	**Reduction of dimension in supervised learning. Application to the study of brain activity** **PhD Laurent Vezard** (with Marie Chavent, Frederique Faïta) **Internships : Julien Clauzel, Nidal El Yacoubi and Emilie Drouineau**
European Project IRSES FP7 MARIE CURIE 2013-2016	**Analysis and Classification of mental states of vigilance with evolutionary computation** **PhD Yuliana Martinez, PhD Enrique Naredo, PhD Emigdio Flores,** **Internships : Victor Lopez Lopez, Uriel Lopez Islas, Enrique Hernandez and Luis Herrera**
HUMO Micro-projects GIS ALBATROS 2015-2017	**Human monitoring (x 3)** 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 GIS ALBATROS 2017	**Apport de l'analyse temps-frequence pour l'estimation de de micro-doppler** Eric Grivel. **Internship Sabrina Macchour** Starting point of the **CIFRE PhD of Jimmy Bondu**
FracLab toolbox 1999- ?	**Matlab toolbox for multifractal analysis and signal processing** **https://project.inria.fr/fraclab/**

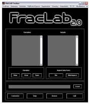
Participation
ANR – RNTS HEVEA
ANR BNPSI
ARC M2A3PC
Micro-Projet GIS Albatros x 2
...



4 countries, 50 months of mobility
31 journal papers, 28 proceedings,
2 books, 1 book chapter
7 PhD defenses



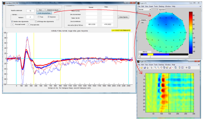




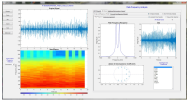
Management of research projects

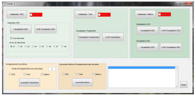
PSI – CNRS 2008-2011	**Psychology and sound interactions** Fréderique Faïta, Joseph Larralde, Pierre-Henry Vulliard, Myriam Desainte-Catherine. **Internship Mathieu Carpentier**
PSI – REGION 2010-2014	**Reduction of dimension in supervised learning. Application to the study of brain activity** **PhD Laurent Vezard** (with Marie Chavent, Frederique Faïta) **Internships : Julien Clauzel, Nidal El Yacoubi and Emilie Drouineau**
European Project IRSES FP7 MARIE CURIE 2013-2016	**Analysis and Classification of mental states of vigilance with evolutionary computation** **PhD Yuliana Martinez, PhD Enrique Naredo, PhD Emigdio Flores,** **Internships : Victor Lopez Lopez, Uriel Lopez Islas, Enrique Hernandez and Luis Herrera**
HUMO Micro-projects GIS ALBATROS 2015-2017	**Human monitoring (x 3)** 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 GIS ALBATROS 2017	**Apport de l'analyse temps-frequence pour l'estimation de de micro-doppler** Eric Grivel. **Internship Sabrina Macchour** Starting point of the **CIFRE PhD of Jimmy Bondu**
FracLab toolbox 1999- ?	**Matlab toolbox for multifractal analysis and signal processing** **https://project.inria.fr/fraclab/**

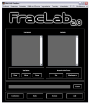
Participation
ANR – RNTS HEVEA
ANR BNPSI
ARC M2A3PC
Micro-Projet GIS Albatros x 2
...

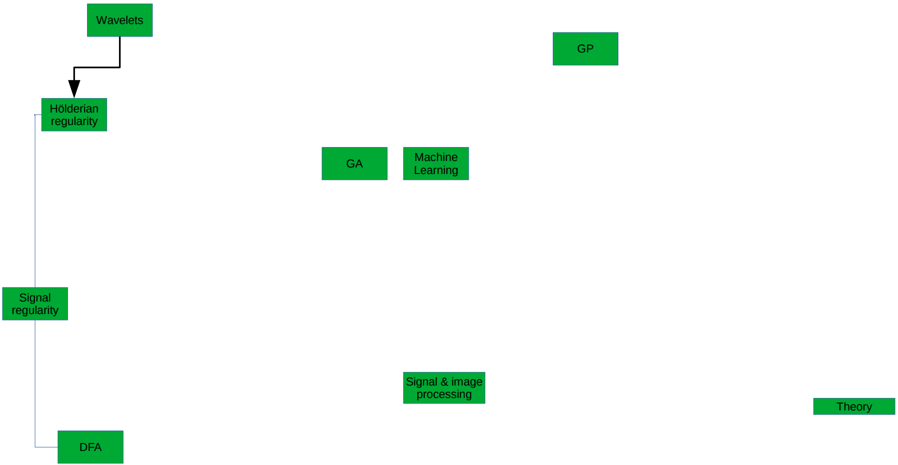


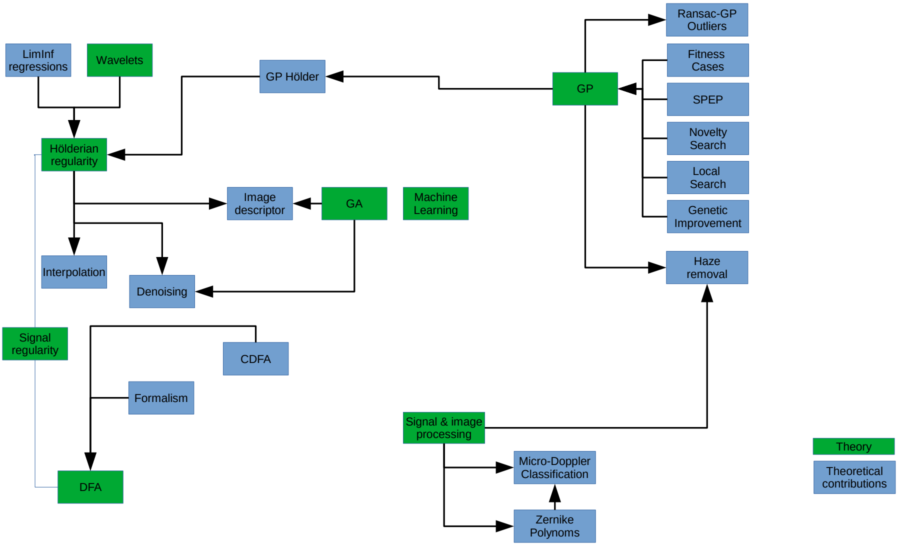
4 countries, 50 months of mobility
31 journal papers, 28 proceedings,
2 books, 1 book chapter
7 PhD defenses

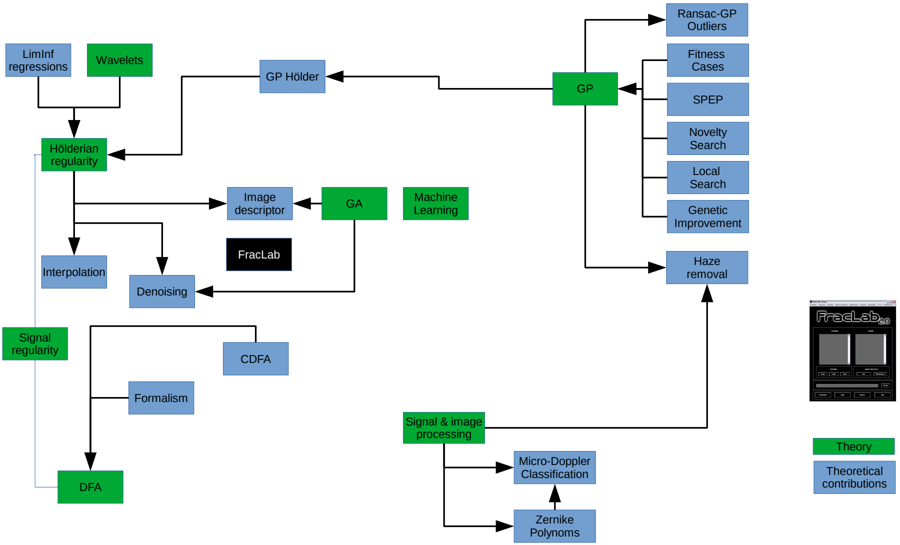


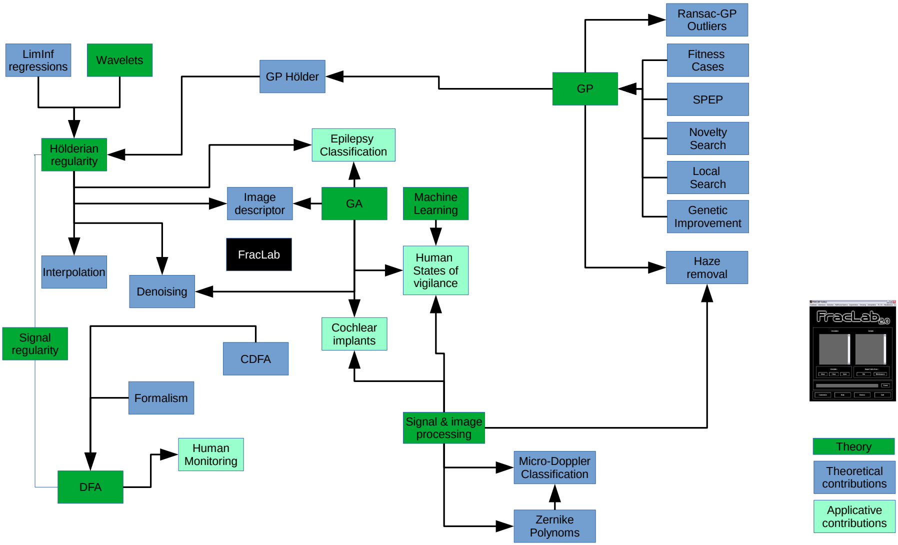




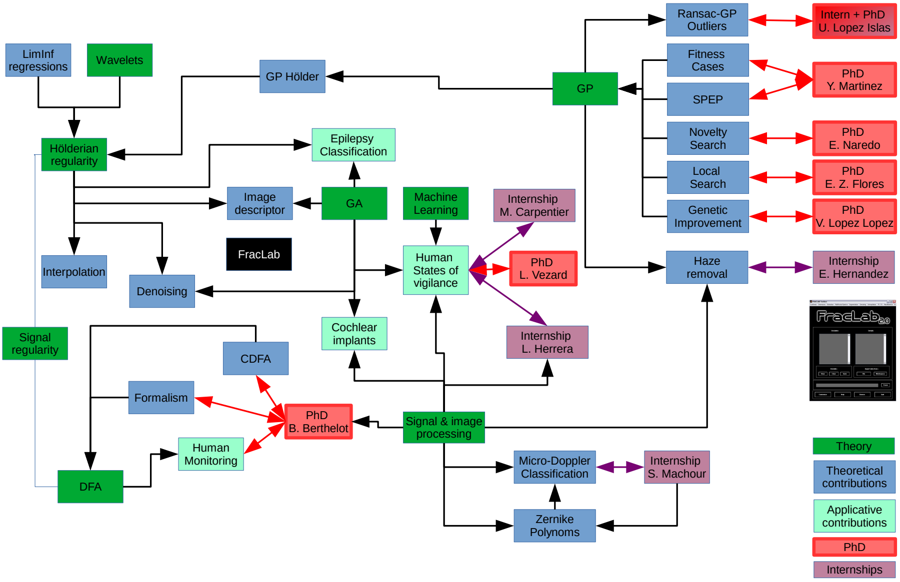

Wavelets
GP
Hölderian regularity
GA
Machine Learning
Signal regularity
Signal & image processing
Theory
DFA


Ransac-GP Outliers
LimInf regressions
Wavelets
GP Hölder
Fitness Cases
GP
SPEP
Hölderian regularity
Novelty Search
Local Search
Image descriptor
GA
Machine Learning
Genetic Improvement
Interpolation
Haze removal
Denoising
Signal regularity
CDFA
Formalism
Signal & image processing
Micro-Doppler Classification
Theory
Theoretical contributions
DFA
Zernike Polynoms

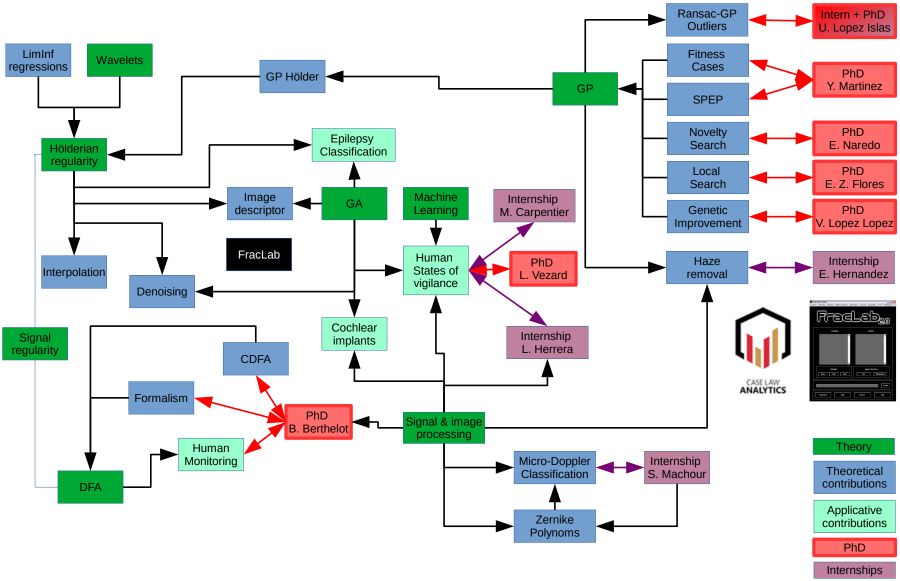

Ransac-GP Outliers
LimInf regressions
Wavelets
GP Hölder
Fitness Cases
GP
SPEP
Hölderian regularity
Novelty Search
Local Search
Image descriptor
GA
Machine Learning
Genetic Improvement
FracLab
Haze removal
Interpolation
Denoising
Signal regularity
CDFA
Formalism
Signal & image processing
Micro-Doppler Classification
DFA
Zernike Polynoms
Theory
Theoretical contributions



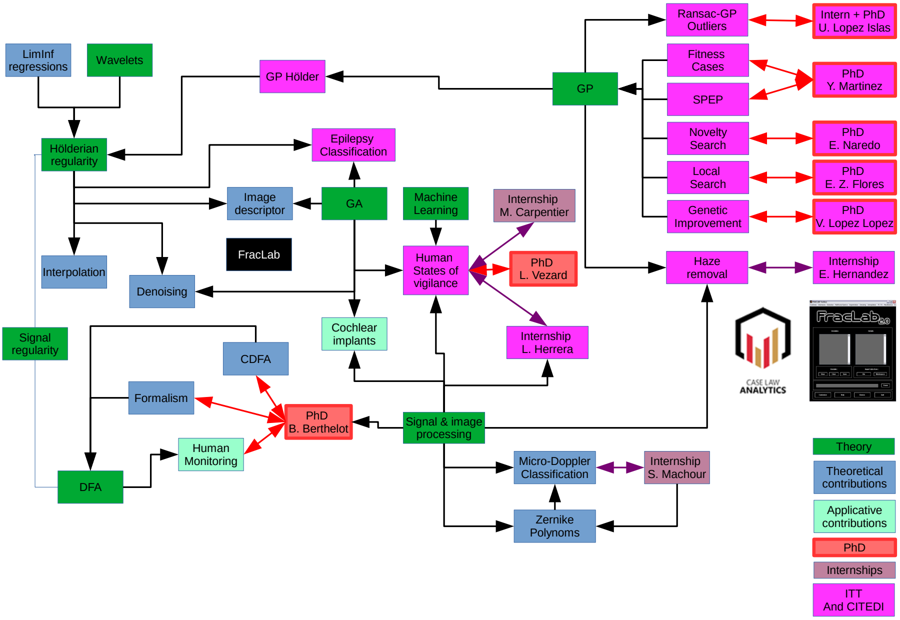
LimInf regressions
Wavelets
GP Hölder
Ransac-GP Outliers
Fitness Cases
GP
SPEP
Hölderian regularity
Epilepsy Classification
Novelty Search
Local Search
Image descriptor
GA
Machine Learning
Genetic Improvement
FracLab
Interpolation
Denoising
Human States of vigilance
Haze removal
Cochlear implants
Signal regularity
CDFA
Formalism
Signal & image processing
Human Monitoring
Micro-Doppler Classification
DFA
Zernike Polynoms
Theory
Theoretical contributions
Applicative contributions



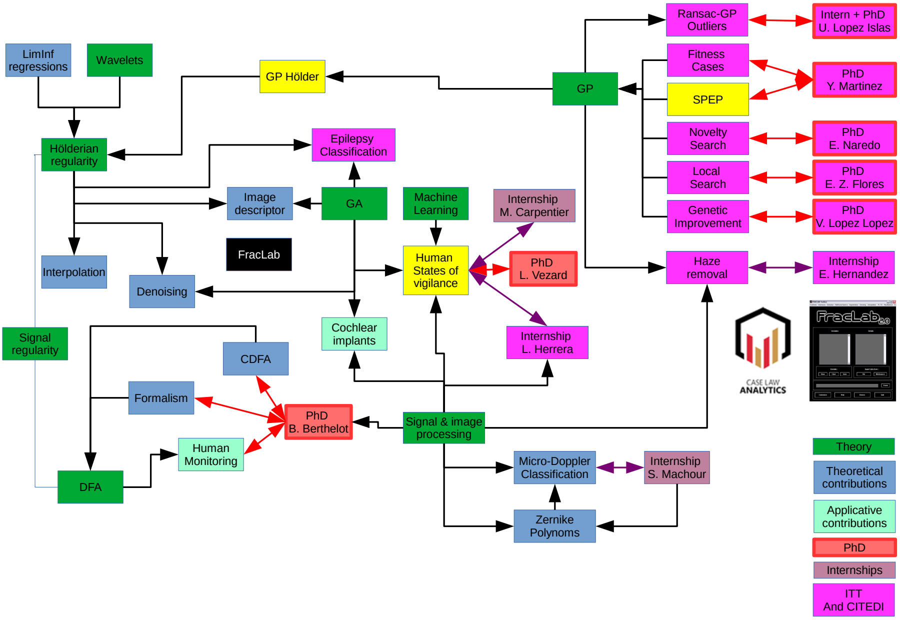
Ransac-GP Outliers
Intern + PhD U. Lopez Islas
LimInf regressions
Wavelets
GP Hölder
GP
Fitness Cases
SPEP
PhD Y. Martinez
Novelty Search
PhD E. Naredo
Local Search
PhD E. Z. Flores
Genetic Improvement
PhD V. Lopez Lopez
Hölderian regularity
Epilepsy Classification
Image descriptor
GA
Machine Learning
Internship M. Carpentier
FracLab
Interpolation
Denoising
Human States of vigilance
PhD L. Vezard
Haze removal
Internship E. Hernandez
Cochlear implants
Internship L. Herrera
Signal regularity
CDFA
Formalism
PhD B. Berthelot
Signal & image processing
Human Monitoring
DFA
Micro-Doppler Classification
Internship S. Machour
Zernike Polynoms
Theory
Theoretical contributions
Applicative contributions
PhD
Internships



LimInf regressions
Wavelets
GP Hölder
Hölderian regularity
Epilepsy Classification
Image descriptor
GA
Machine Learning
Internship M. Carpentier
Interpolation
Denoising
FracLab
Human States of vigilance
PhD L. Vezard
Internship L. Herrera
Cochlear implants
Signal regularity
CDFA
Formalism
PhD B. Berthelot
Signal & image processing
Human Monitoring
DFA
Micro-Doppler Classification
Internship S. Machour
Zernike Polynoms
GP
Ransac-GP Outliers
Intern + PhD U. Lopez Islas
Fitness Cases
SPEP
PhD Y. Martinez
Novelty Search
PhD E. Naredo
Local Search
PhD E. Z. Flores
Genetic Improvement
PhD V. Lopez Lopez
Haze removal
Internship E. Hernandez
CASE LAW ANALYTICS
FracLab2.0
Theory
Theoretical contributions
Applicative contributions
PhD
Internships



LimInf regressions
Wavelets
GP Hölder
GP
Ransac-GP Outliers
Intern + PhD U. Lopez Islas
Fitness Cases
SPEP
PhD Y. Martinez
Novelty Search
PhD E. Naredo
Local Search
PhD E. Z. Flores
Genetic Improvement
PhD V. Lopez Lopez
Hölderian regularity
Epilepsy Classification
Image descriptor
GA
Machine Learning
Internship M. Carpentier
FracLab
Human States of vigilance
PhD L. Vezard
Haze removal
Internship E. Hernandez
Interpolation
Denoising
Cochlear implants
Internship L. Herrera
Signal regularity
CDFA
Formalism
PhD B. Berthelot
Signal & image processing
Human Monitoring
DFA
Micro-Doppler Classification
Internship S. Machour
Zernike Polynoms
CASE LAW ANALYTICS
FracLab2.0
Theory
Theoretical contributions
Applicative contributions
PhD
Internships
ITT And CITEDI



LimInf regressions
Wavelets
GP Hölder
GP
Ransac-GP Outliers
Intern + PhD U. Lopez Islas
Fitness Cases
SPEP
PhD Y. Martinez
Novelty Search
PhD E. Naredo
Local Search
PhD E. Z. Flores
Genetic Improvement
PhD V. Lopez Lopez
Hölderian regularity
Epilepsy Classification
Image descriptor
GA
Machine Learning
Internship M. Carpentier
FracLab
Interpolation
Denoising
Human States of vigilance
PhD L. Vezard
Haze removal
Internship E. Hernandez
Cochlear implants
Internship L. Herrera
Signal regularity
CDFA
Formalism
PhD B. Berthelot
Signal & image processing
Human Monitoring
DFA
Micro-Doppler Classification
Internship S. Machour
Zernike Polynoms
CASE LAW ANALYTICS
FracLab 2.0
Theory
Theoretical contributions
Applicative contributions
PhD
Internships
ITT And CITEDI

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

PART 1: Artificial Evolution

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

Artificial Darwinism

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

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

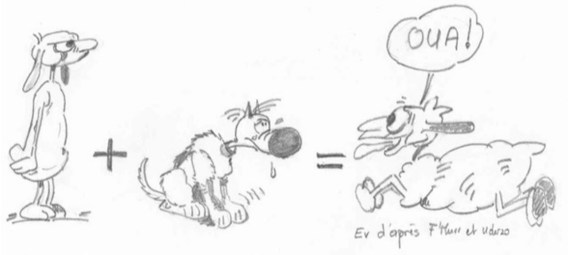
A set of techniques grouped under a generic term

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

Ingredients

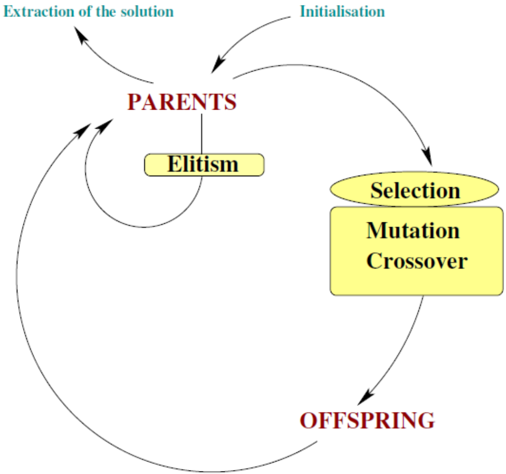
Selection

Population

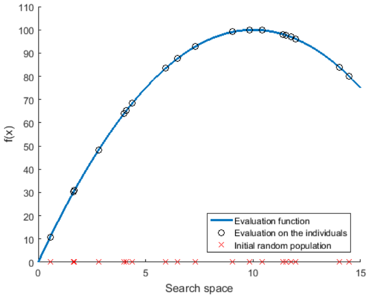


Genetic Operators

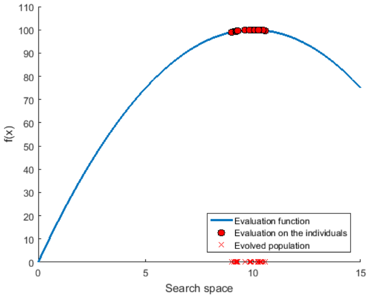
Evolutionary loop



Potential Solutions = Individuals in a population



Initial random population



Evolved population

Artificial Darwinism

- 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

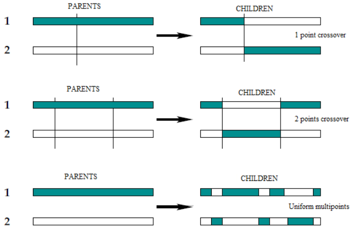
Genetic Programming

- Functional representation: Genetic programming
- Example: Using GP for regression

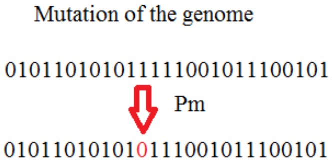
Discrete representation: Genetic Algorithms

Each individual is represented by a binary string.

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



Crossover



Mutation

Continuous representation: Evolution Strategies

Each individual is a vector in R^n.

Hans-Paul Schwefel (1970)

Barycentric crossover

$\forall i \in \{1, .., n\}, x_i^{children} = \alpha x_i^{father} + (1 - \alpha) x_i^{mother}$

α random value in $[-\epsilon, 1 + \epsilon]$.

Gaussian mutation

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

Two parameters P_m and σ.

Continuous representation: Evolution Strategies

Each individual is a vector in R^n.

Hans-Paul Schwefel (1970)

Barycentric crossover

$\forall i \in \{1, .., n\}, x_i^{children} = \alpha x_i^{father} + (1 - \alpha) x_i^{mother}$

α random value in $[-\epsilon, 1 + \epsilon]$.

Gaussian mutation

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

Two parameters P_m and σ.

Demo

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

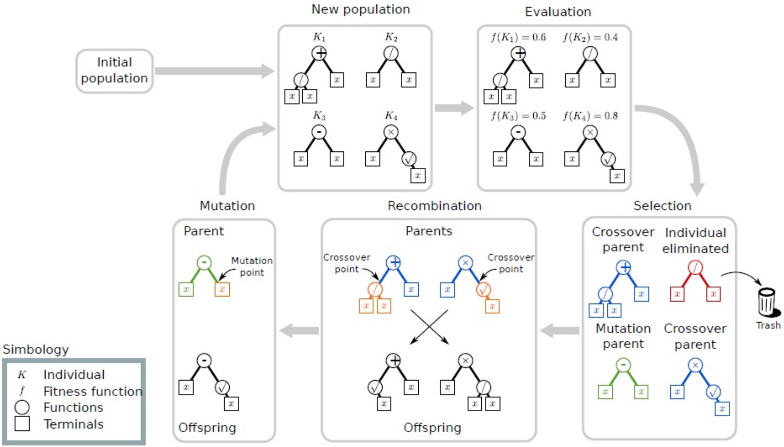
Genetic Algorithms
- Discrete representation: Genetic Algorithms

Evolution strategies
- Continuous representation: Evolution Strategies

Genetic Programming
- Functional representation: Genetic programming
- Example: Using GP for regression

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].

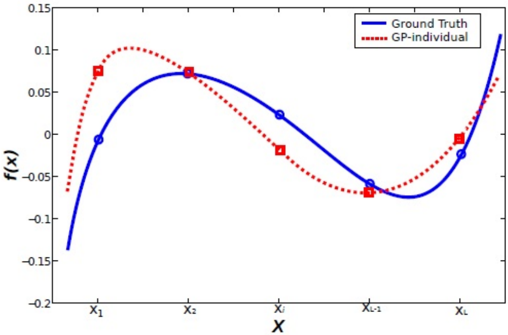


Example: Using GP for regression

Symbolic Regression

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

	X	Y
1	0.2	0.00
2	0.4	0.06
..	...	...
N	0.9	-0.03



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

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).

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

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

1. **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.

1. **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.

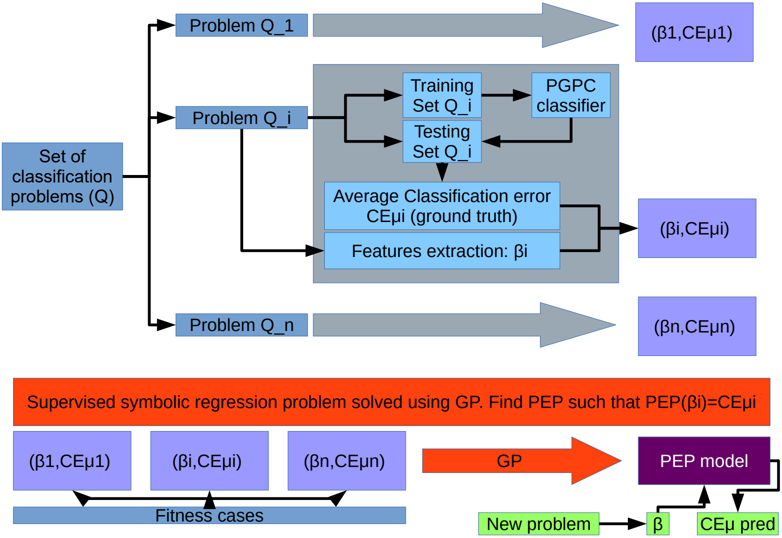
   [Graff and Poli, 2010, Graff and Poli, 2011, Graff et al., 2013, Trujillo et al., 2011a, Trujillo et al., 2011b, Trujillo et al., 2011c].

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.

Proposal



Set of classification problems (Q)
Problem Q_1
Problem Q_i
Problem Q_n
(β1,CEμ1)
Training Set Q_i
PGPC classifier
Testing Set Q_i
Average Classification error CEμi (ground truth)
Features extraction: βi
(βi,CEμi)
(βn,CEμn)
Supervised symbolic regression problem solved using GP. Find PEP such that PEP(βi)=CEμi
(β1,CEμ1)
(βi,CEμi)
(βn,CEμn)
Fitness cases
GP
PEP model
New problem
β
CEμ pred

Geometric mean (SD):

Measures the homogeneity of covariances [Michie1994,So1999].

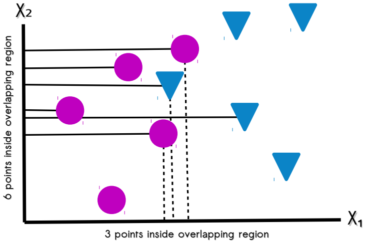
$$SD = exp\left\{\frac{M}{m\sum_{i=1}^{C}(n_i-1)}\right\}$$

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. $n=\sum_{i=1}^{C} n_i$.

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

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

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



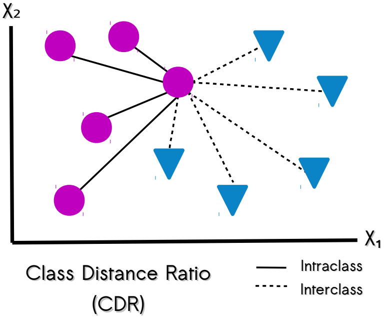
Feature Efficiency (FE)

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

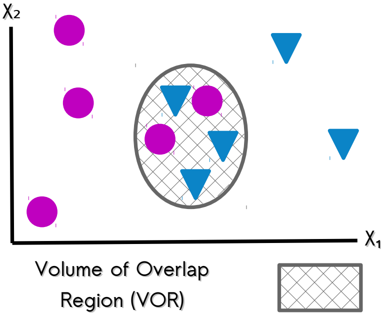
$$FE_j = \left(1 - \frac{\eta_j}{n}\right)$$

where η_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 $[1, m]$.



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):

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.

Feature Extraction

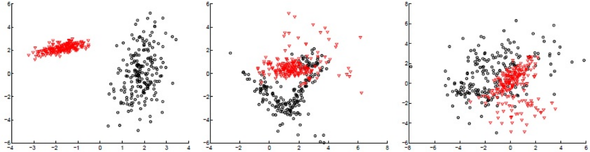
Canberra Distance (CD):

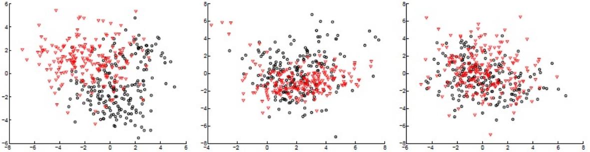
Provides a numerical measure of the distance between pairs of points in a vector space. Suppose a problem has m dimensions, we take a rank statistic of the samples of each class, call it x_i for class 1 and y_i for class 2, for the i-th dimension. This produces two vectors $\mathbf{x}$ and $\mathbf{y}$, 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.

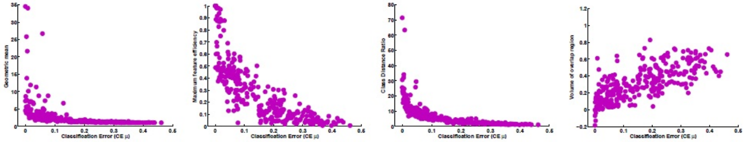
Synthetic problems


(a) $CE\mu = 0\ \sigma = 0$ (b) $CE\mu = 0.14\ \sigma = 0.03$ (c) $CE\mu = 0.17\ \sigma = 0.03$

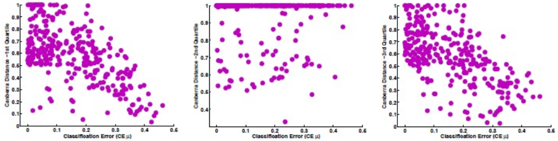

(d) $CE\mu = 0.21\ \sigma = 0.03$ (e) $CE\mu = 0.36\ \sigma = 0.04$ (f) $CE\mu = 0.46\ \sigma = 0.04$

Examples of synthetic classification problems, specifying the CEμ and standard deviation σ achieved by PGPC. These ordered from the lowest CEμ (easiest) to the highest CEμ (hardest).

Correlation



(a) SD: $\rho = -0.42$ (b) FE: $\rho = -0.78$ (c) CDR: $\rho = -0.62$ (d) VOR: $\rho = 0.72$



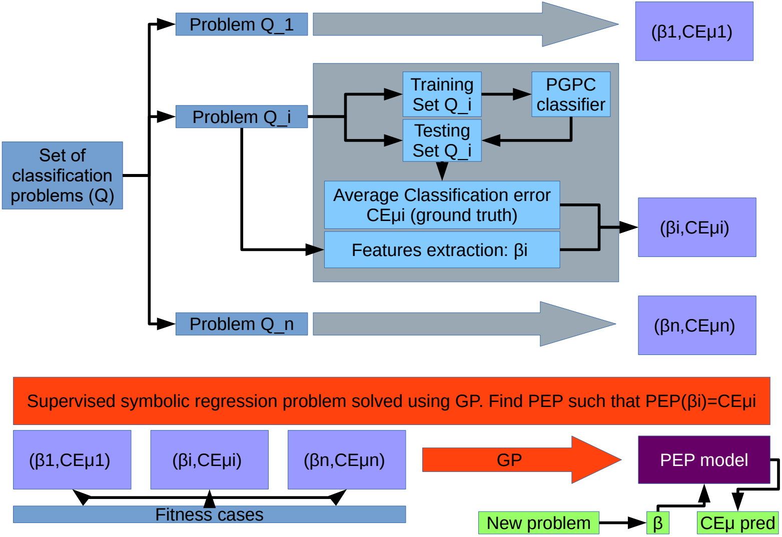
(e) CD-1: $\rho = -0.62$ (f) CD-2: $\rho = -0.03$ (g) CD-3: $\rho = -0.61$

Relationship between the CEμ (x-axis) and each descriptive feature (y-axis) for all problems $p \in \mathcal{Q}$, where ρ specifies Pearson's correlation coefficient.

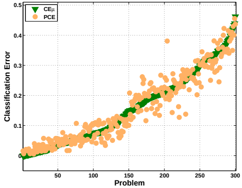
Three versions of the problem are posed, each with a different terminal set defined as subsets of all extracted features(4F, 5F, 7F).

- 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.

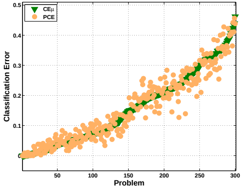
Building PEP model



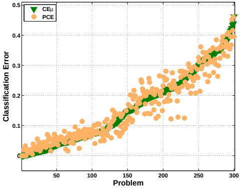
Testing PEP models in synthetic classification problems



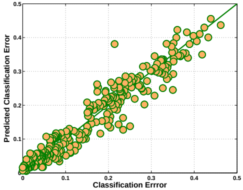
(a) [*PEP-4F:* RMSE = 0.0318]



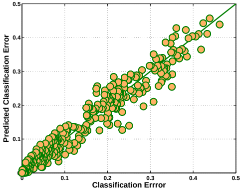
(b) [*PEP-5F:* RMSE = 0.0295]



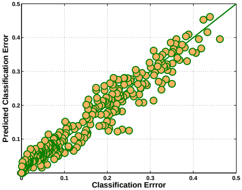
(c) [*PEP-7F:* RMSE = 0.0317]



(d) [*PEP-4F:* $\rho = 0.9634$]



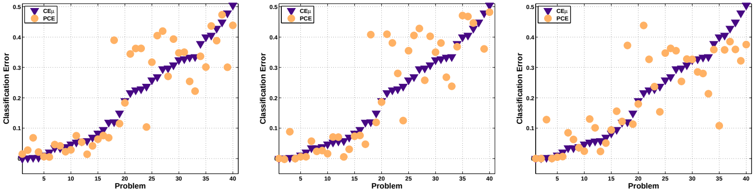
(e) [*PEP-5F:* $\rho = 0.9688$]



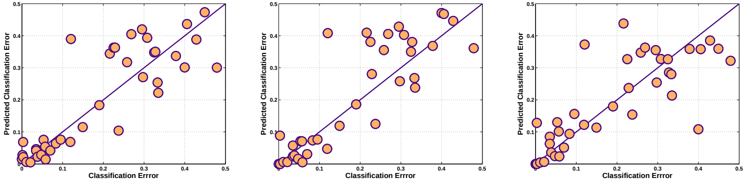
(f) [*PEP-7F:* $\rho = 0.9636$]

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 CEμ. Second line: scatter plots of the PCE and the CEμ.

Testing PEP models in real-world classification problems



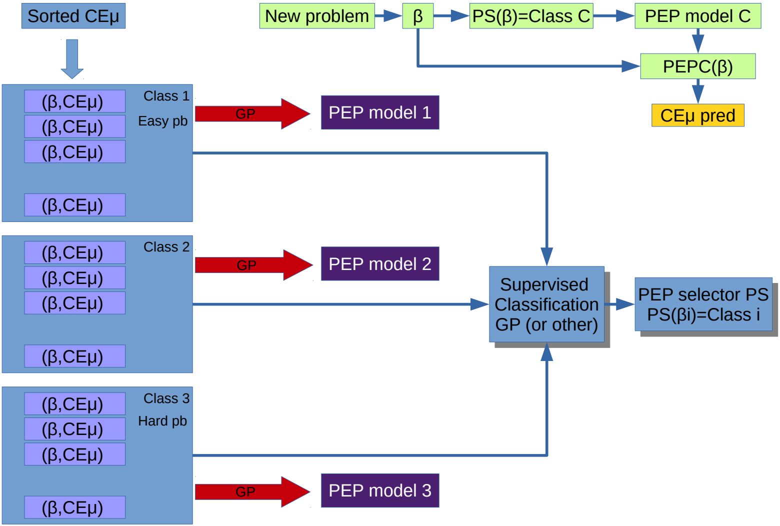
(a) [*PEP-4F:* RMSE = 0.0828] (b) [*PEP-5F:* RMSE = 0.0929] (c) [*PEP-7F:* RMSE =0.0930]



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

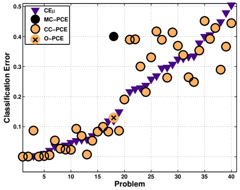
Performance prediction of the best PEP models evolved with the different feature set. PEP-4F (left), PEP-5F (middle) and PEP-7F (right). First line: PCE of the best solution and the know CEμ. Second line: scatter plots of the PCE and the CEμ.

Proposal

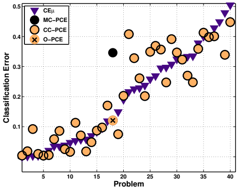


Sorted CEμ
(β,CEμ)
Class 1
Easy pb
GP
PEP model 1
Class 2
GP
PEP model 2
Class 3
Hard pb
GP
PEP model 3
Supervised Classification GP (or other)
PEP selector PS PS(βi)=Class i
New problem
β
PS(β)=Class C
PEP model C
PEPC(β)
CEμ pred

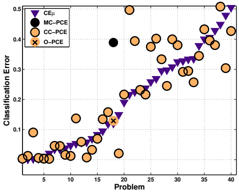
SPEP using two groups



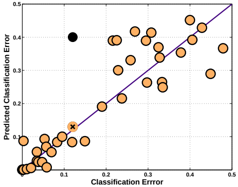
(a) [*4F:* RMSE = 0.0818]



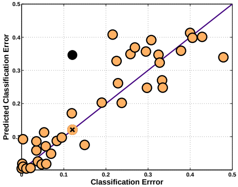
(b) [*5F:* RMSE = 0.0736]



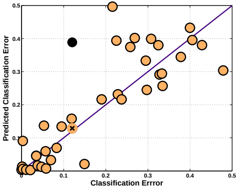
(c) [*7F:* RMSE = 0.0897]



(d) [*4F:* $\rho = 0.8717$]



(e) [*5F:* $\rho = 0.8981$]



(f) [*7F:* $\rho = 0.8514$]

Performance prediction of the best Ensemble-2 solutions for each feature set: 4F (left), 5F (middle) and 7F (right). First line: ground truth CEμ of each problem (triangles) and the corresponding PCE (circles). Second line: scatter plots between the CEμ 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).

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

- 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.

- 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.

- 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

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 signal

Signal Regularity

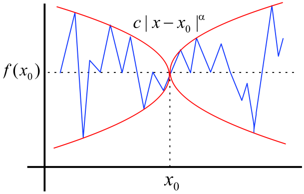
The Hölder pointwise exponent is the most common tool used to measure the regularity of a signal at a given point.

Definition

Let f be a function from $\mathbb{R}$ to $\mathbb{R}$, $s > 0$, $s \in \mathbb{R} \backslash \mathbb{N}$ and $x_0 \in \mathbb{R}$. 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

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

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 signal

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,

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

In terms of oscillations, this condition can be written:
A function $f(t)$ is Hölderian of exponent α at t, with $0 < \alpha < 1$ if there is a constant c such that for any τ,

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

with

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

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.**

Regression of wavelet coefficients

Theorem

(S. Jaffard)

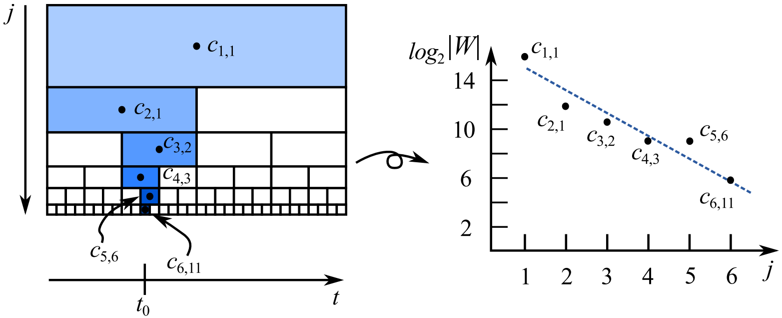
$$|c_{j,k}| \leq \; c2^{-j(\alpha+\frac{1}{2})}(1+|2^j t_0 - k|)^{\alpha} \quad \forall j,k \in \mathbb{Z}^2$$

Conversely ;

$$If \quad \forall j,k \in \mathbb{Z}^2 \;\; one \;\; has \;\; |c_{j,k}| \leq \; c2^{-j(\alpha+\frac{1}{2})}(1+|2^j t_0 - k|)^{\alpha'}$$

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

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.

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.

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:

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

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.

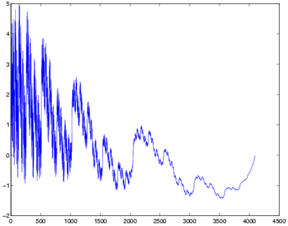
We note k but the value is $\lfloor \frac{t_0+1}{2^{n-j+1}} \rfloor$.

Definitions
Signal Regularity

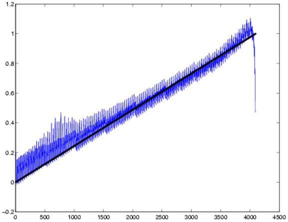
Estimation
Oscillations
Regression of wavelet coefficients

Application
Estimation on synthetic signal

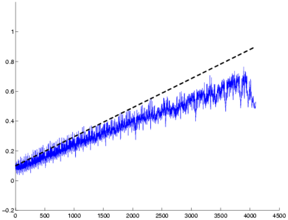
Estimation on synthetic signal



GWei



WCR



OSC

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-Aïnseba and Leonardo Trujillo.

Goal

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

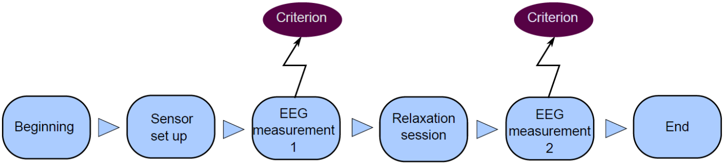
EEG data Acquisition
- Acquisition Protocole

Feature Extraction
- Slope Criterion

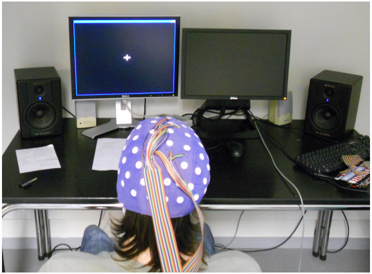
Evolutionary Algorithm
- Design
- Results

Acquisition Protocole

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



Acquisition Protocole



- 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

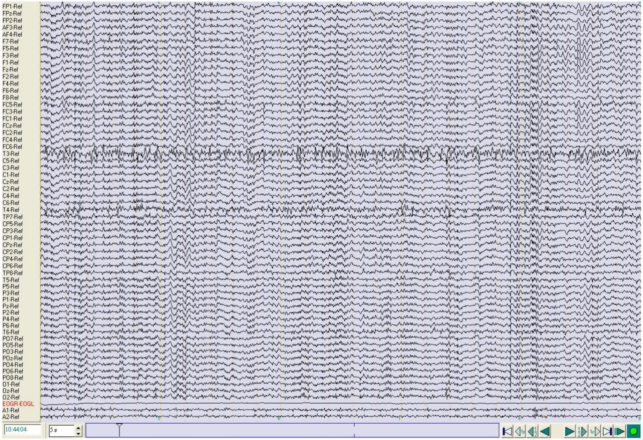
Relaxation session

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).

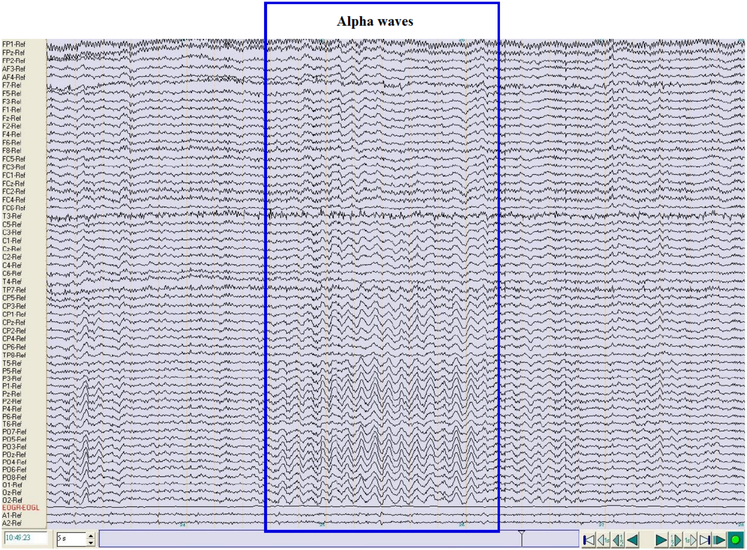
Acquisition Protocole

3 minutes of EEG recording before relaxation.

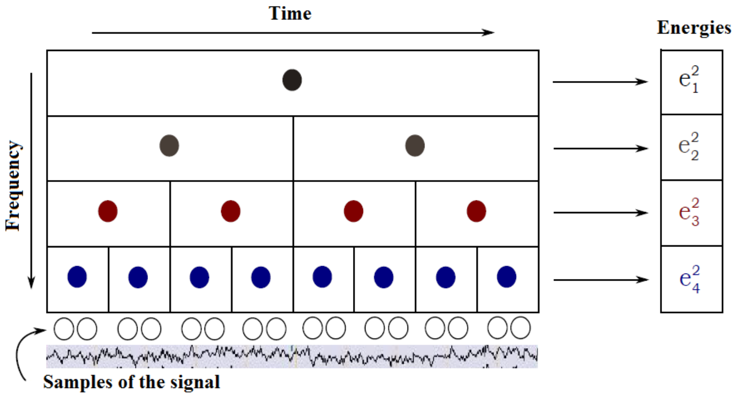


Acquisition Protocole

3 minutes of EEG recording after relaxation.

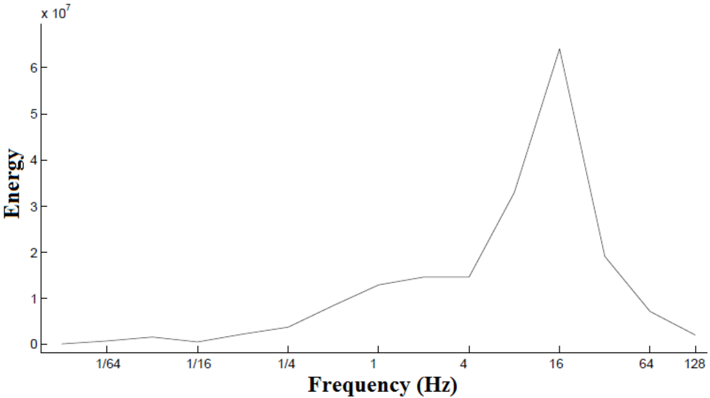


The dyadic grid gives a spatio-frequential representation of the discrete dyadic wavelet decomposition



Slope Criterion

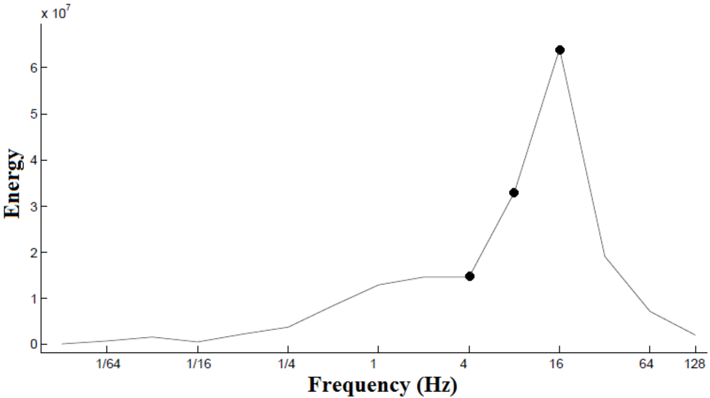
- Alpha: $8 - 12Hz$.
- Waves characteristics of a relaxed state.



Linear regression between 4 and $16Hz$.

Slope Criterion

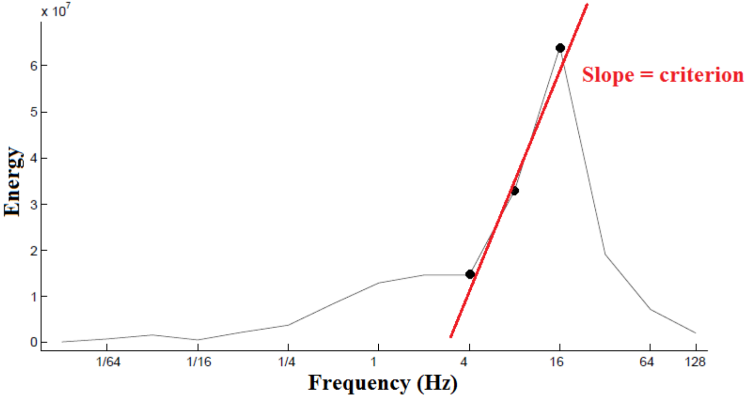
- Alpha: $8 - 12Hz$.
- Waves characteristics of a relaxed state.



Linear regression between 4 and $16Hz$.

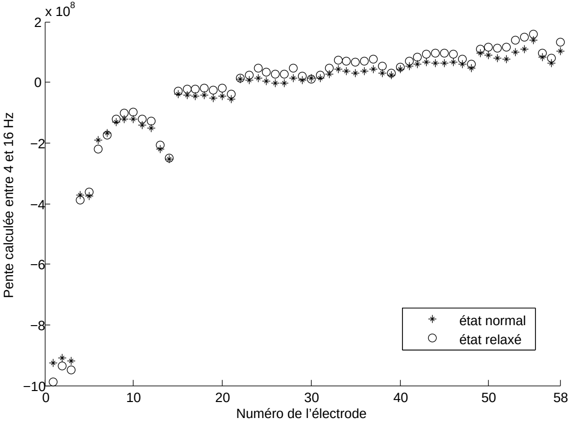
Slope Criterion

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



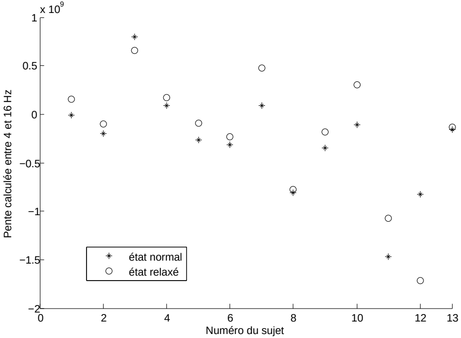
Linear regression between 4 and $16Hz$.

Slope Criterion



Slope criterion, sum on subjects for each electrode

Slope Criterion



Slope criterion, sum on electrodes for each subjects

$\Rightarrow$ Very strong inter-subject variability

$\Rightarrow$ This criterion that does not allow to build a powerful classifier for different subjects.

Classification: Usual methods

	K nearest neighbors	Binary decision trees	Random forests	Discriminant PLS	Sparse Discriminant 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.

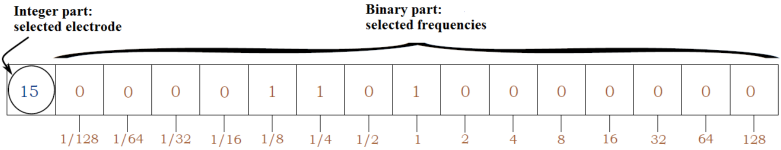
This approach is not efficient

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

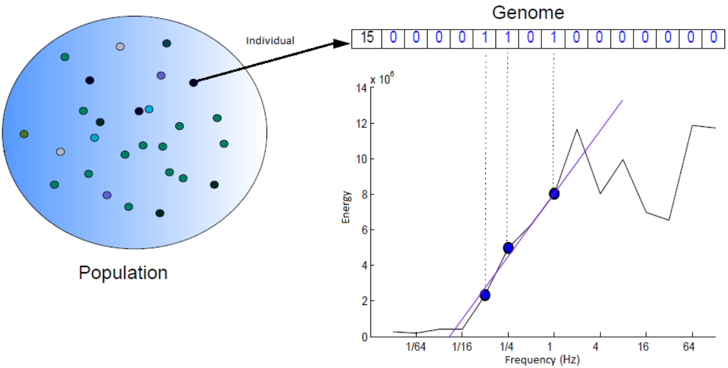
$\Rightarrow$ **Find the relevant electrodes.**

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

Example of a genome in the evolutionary algorithm



Relationship between the genome and the calculation of the slope criterion



Average correct classification rate

Evaluation Method	CCR	
	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.

Best genome

Evaluation method	BEST genome		
	Selected electrode	Selected frequency (Hz)	Correct classification 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

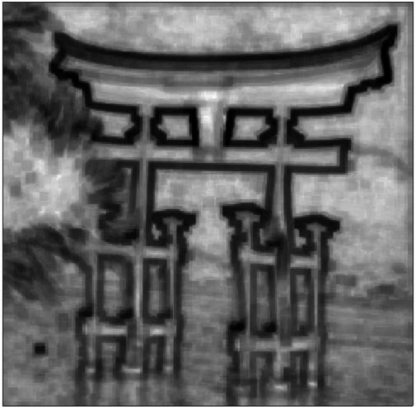
Contribution

Training set

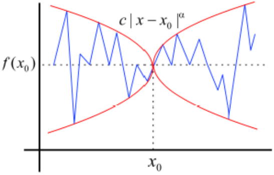
Results

Hölder exponent

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



General motivation

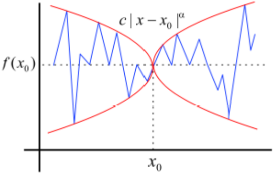


$$\alpha_p(x_0) = \liminf_{h\to 0} \frac{\log|f(x_0+h) - f(x_0)|}{\log|h|}$$

Hölderian envelope of signal f at point x_0

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

General motivation

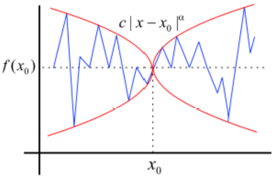


$$\alpha_p(x_0) = \liminf_{h \to 0} \frac{\log|f(x_0+h) - f(x_0)|}{\log|h|}$$

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**;

General motivation



Hölderian envelope of signal f at point x_0

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

- 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

Contribution

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

Contribution

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

Contribution

- 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.

Contribution

- 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.

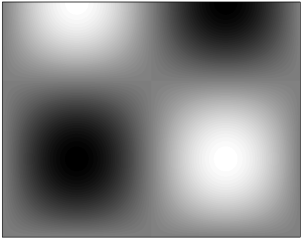
We generate three groups of images with **FracLab**, using three different functions that take as input the point coordinates (x, y) of an image and provide as output the desired regularity; these functions are:

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

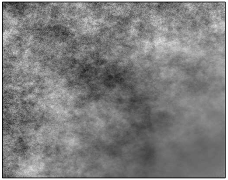
(a) Polynomial p_1



(b) Sine p_2

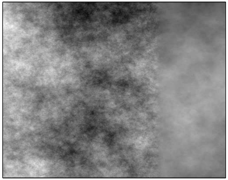
(c) Exponential p_3

Training set

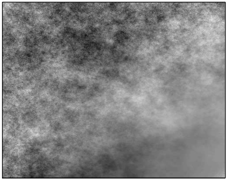


(a) Polynomial

(b) Sine

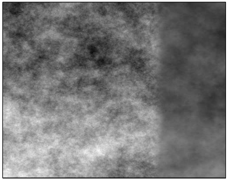


(c) Exponential



(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

Results: Real Images

Original Image

Traditional Method

GP-Estimator

GP-Estimator