Fitness landscape analysis: an overview for single- and multi-objective optimization problems

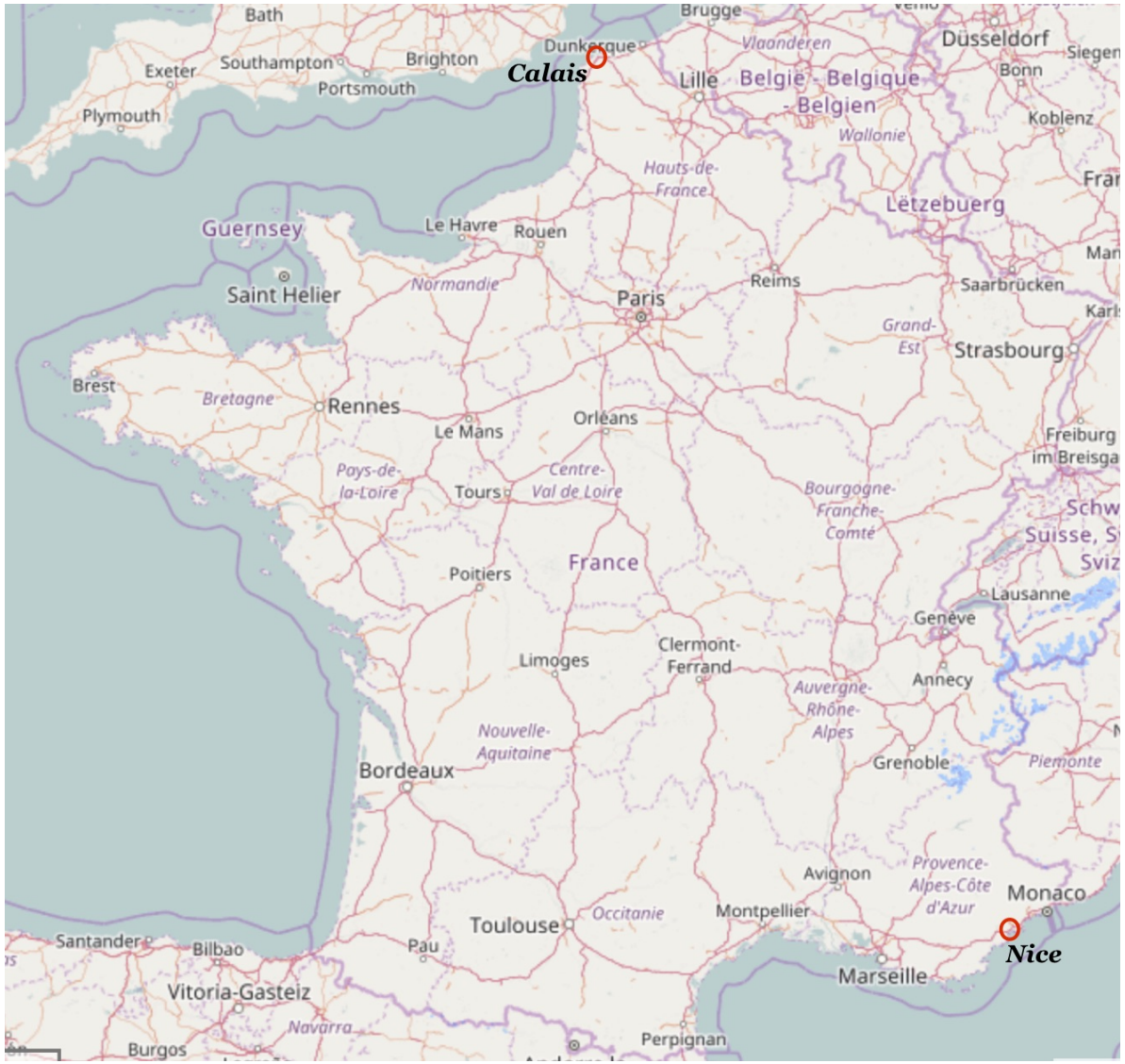
Sébastien Verel

LISIC - Université du Littoral Côte d'Opale, Calais, France
http://www-lisic.univ-littoral.fr/~verel/

University Exeter

18th, May, 2017

Personal situation map



source : OpenStreetMap

Situation map : Université du Littoral Côte d'Opale

LISIC : laboratory of computer science, signal, image

Boulogne (fishing port), Calais (transportation port),
Dunkerque (industrial port), Saint Omer (glass industry).

⇝ Lille (0h30), London (1h), Bruxelles (1h15), Paris (1h30), Amsterdam (2h)

Outline

- Short introduction on fitness landscape
  mutlimodality, ruggedness
- Single-objective optimization :
  Local Optima Network
- Multi-objective optimization :
  Features of fitness landscape

Mono-objective Optimization

- Search space : set of candidate solutions

$$X$$

- Objective fonction : quality criterion (or non-quality)

$$f : X \to \mathbb{R}$$

X discrete : combinatorial optimization
$X \subset \mathbb{R}^n$: numerical optimization

Solve an optimization problem (maximization)

$$X^\star = \mathrm{argmax}_X \ f$$

or find an approximation of $X^\star$.

Context : black-box optimization

$x \longrightarrow$  $\longrightarrow f(x)$

No information on the objective definition function f

Objective fonction :

- can be irregular, non continuous, non differentiable, etc.
- given by a computation or a simulation

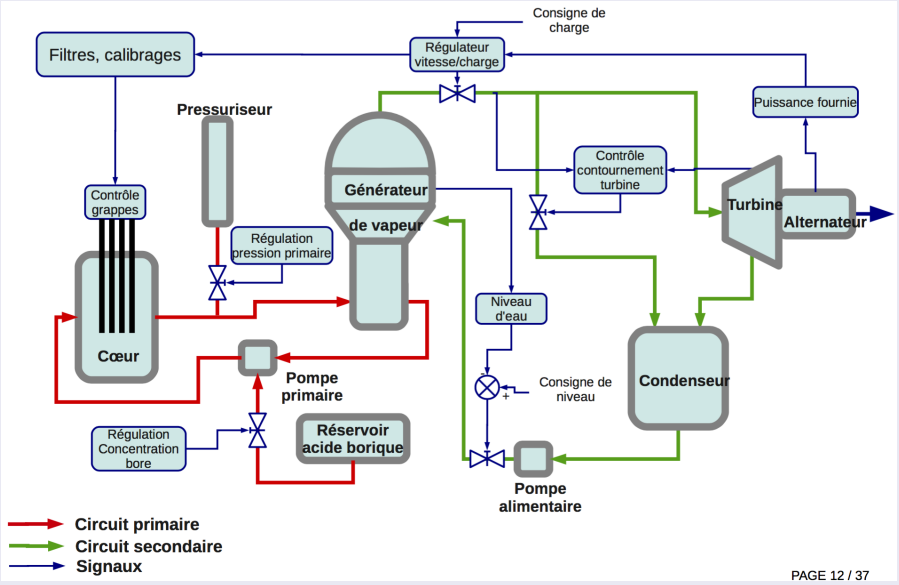
Real-world black-box optimization : first example

PhD of Mathieu Muniglia, Saclay Nuclear Research Centre (CEA), Paris

$x \longrightarrow$

$\longrightarrow f(x)$

$(73, \ldots, 8) \longrightarrow$



Multi-physic simulator

$\longrightarrow \Delta_z P$

Fitness landscape : original plots of S. Wright [Wri32]

S. Wright. "The roles of mutation, inbreeding, crossbreeding, and selection in evolution.", 1932.

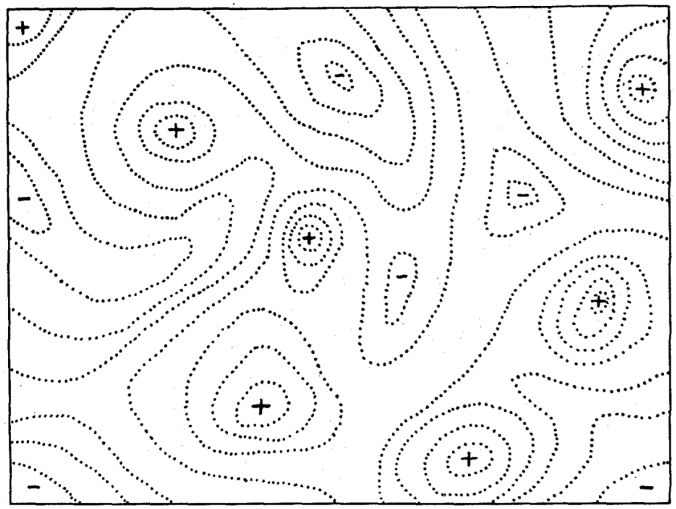


FIGURE 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

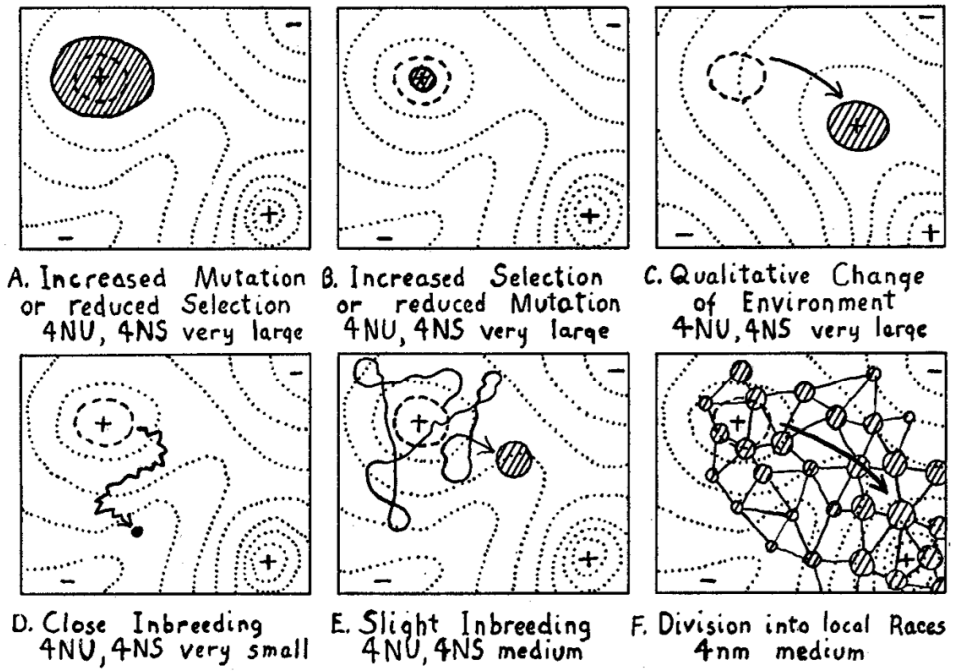
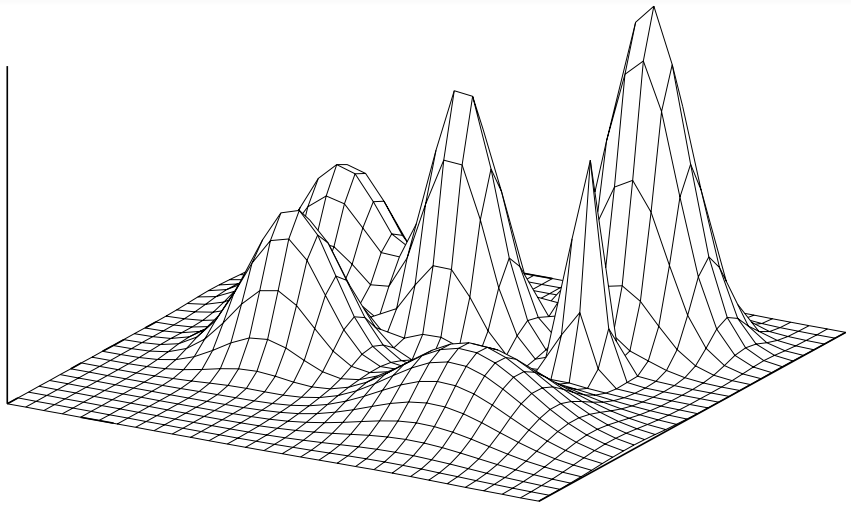


FIGURE 4.—Field of gene combinations occupied by a population within the general field of possible combinations. Type of history under specified conditions indicated by relation to initial field (heavy broken contour) and arrow.

source : Encyclopædia Britannica Online.

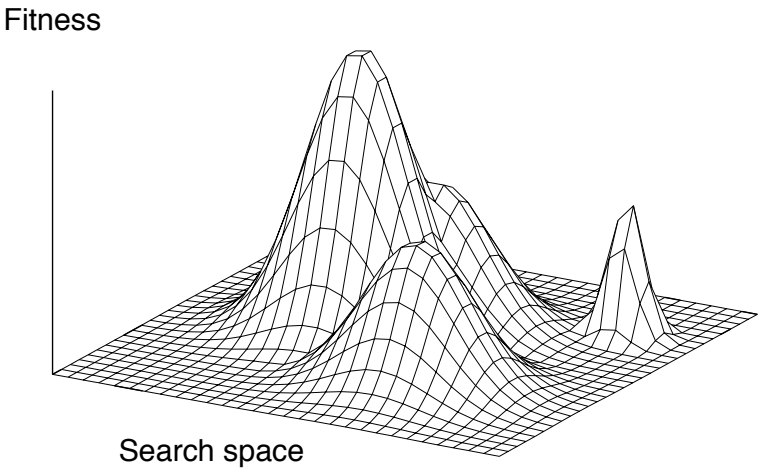
Fitness landscapes in (evolutionary) biology



- Metaphorical uphill struggle across a "fitness landscape"
  - mountain **peaks** represent high "fitness" (ability to survive),
  - **valleys** represent low fitness.
- Evolution proceeds :
  population of organisms
  performs an "**adaptive walk**"

becareful : "2 dimensions instead of many thousands"

Definition of fitness landscape for optimization [Sta02]



Definition

Fitness landscape $(X, \mathcal{N}, f)$:

- **search space** :

$$X$$

- **neighborhood relation** :

$$\mathcal{N} : X \to 2^X$$

- **objective function** :

$$f : X \to \mathbb{R}$$

Fitness landscape analysis

Algebric approach, grey-box :

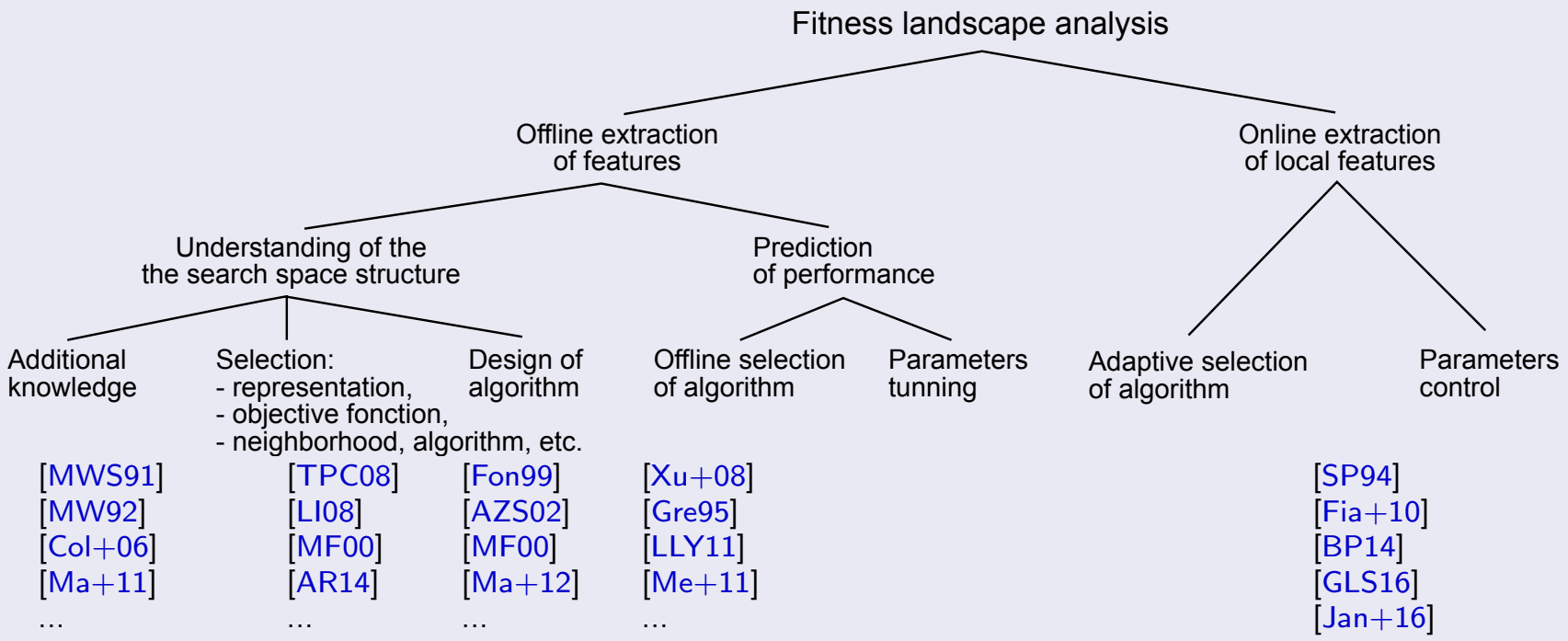
$$\Delta f = \lambda.(f - \bar{f})$$

Statistical approach, black-box :

Problems $\rightsquigarrow$ Features

$\rightsquigarrow$ Algorithm $\rightsquigarrow$ Performances

Goals



Position of fitness landscape analysis

Selection of local search algorithm : Rice framework

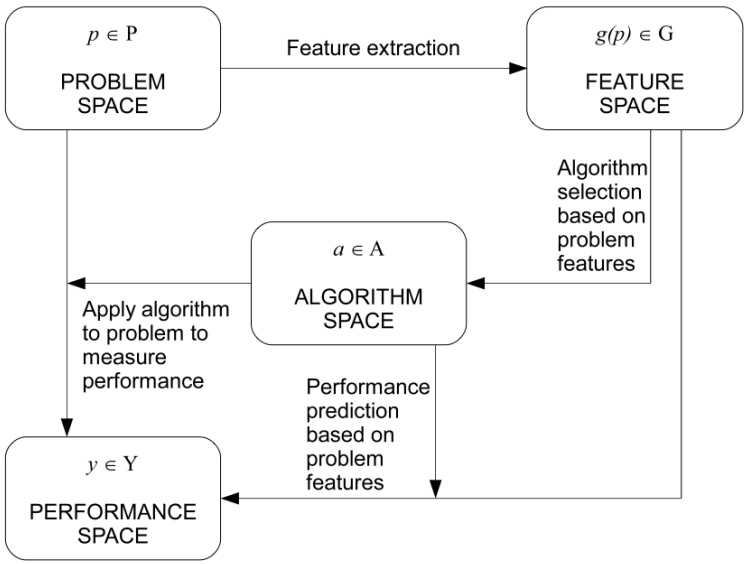


Figure 1.1: A framework for describing the general problems of algorithm selection and performance prediction based on problem features (based Rice's model [132]).

Malan, K. M., Engelbrecht, A. P. (2014). Fitness landscape analysis for metaheuristic performance prediction. In Recent advances in the theory and application of fitness landscapes (pp. 103-132). [ME14]

Position of fitness landscape analysis

Selection of local search algorithm : Rice framework

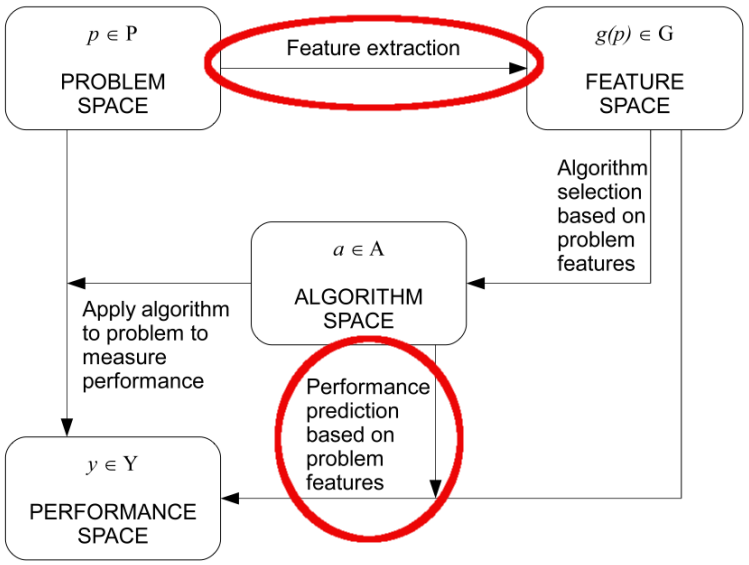


Figure 1.1: A framework for describing the general problems of algorithm selection and performance prediction based on problem features (based Rice's model [132]).

Malan, K. M., Engelbrecht, A. P. (2014). Fitness landscape analysis for metaheuristic performance prediction. In Recent advances in the theory and application of fitness landscapes (pp. 103-132). [ME14]

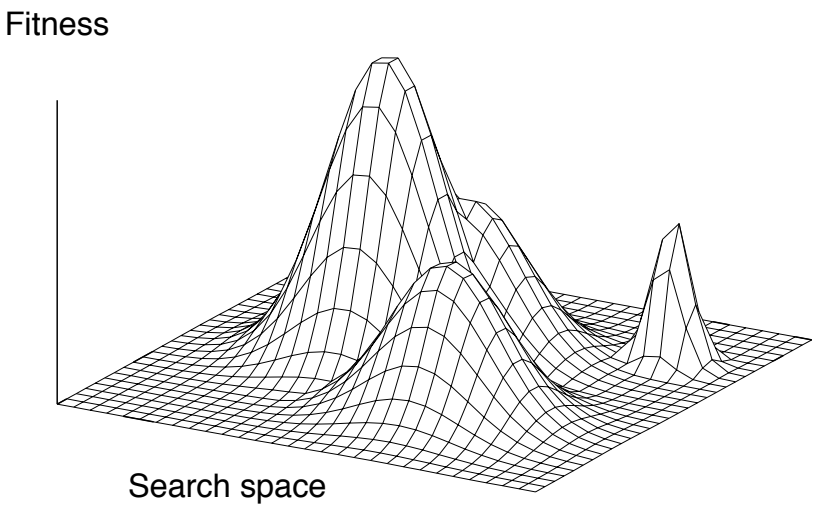
Fitness landscape analysis : features extraction *vs.* performance

Multimodal Fitness landscapes

Local optima x^*

no neighbor solution with strictly higher fitness value (maximization)

$$\forall x \in \mathcal{N}(x^*), \quad f(x) \leqslant f(x^*)$$



Sampling local optima

Basic estimator (Alyahya, K., & Rowe, J. E. 2016 [AR16])

Expected proportion of local optima :

Proportion of local optima in a sample of random solutions

- Complexity : $n \times |\mathcal{N}|$
- Pros :
  unbiased estimator, strong and easy statistical tools
- Cons :
  "resolution rate" is $1/n$

Sampling local optima by adaptive walks

Adaptive walk

$(x_1, x_2, \ldots, x_\ell)$ such that $x_{i+1} \in \mathcal{N}(x_i)$ and $f(x_i) < f(x_{i+1})$

Sampling local optima by adaptive walks

Adaptive walk

$(x_1, x_2, \ldots, x_\ell)$ such that $x_{i+1} \in \mathcal{N}(x_i)$ and $f(x_i) < f(x_{i+1})$

Hill-Climbing algorithm (first-improvement)

Choose initial solution $x \in X$
repeat
  choose $x^{'} \in \{y \in \mathcal{N}(x) \mid f(y) > f(x)\}$
  if $f(x) < f(x^{'})$ **then**
    $x \leftarrow x^{'}$
  end if
until x is a Local Optimum

Sampling local optima by adaptive walks

Adaptive walk

$$(x_1, x_2, \ldots, x_\ell) \text{ such that } x_{i+1} \in \mathcal{N}(x_i) \text{ and } f(x_i) < f(x_{i+1})$$

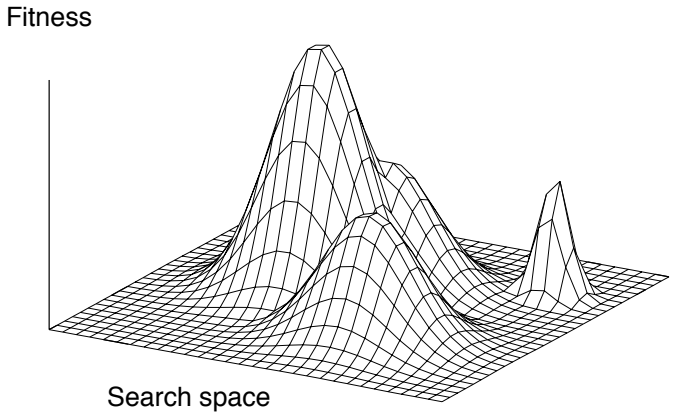
Hill-Climbing algorithm (first-improvement)

Choose initial solution $x \in X$
repeat
  choose $x^{'} \in \{y \in \mathcal{N}(x) \mid f(y) > f(x)\}$
  if $f(x) < f(x^{'})$ **then**
    $x \leftarrow x^{'}$
  end if
until x is a Local Optimum

Basin of attraction of x^*

$$\{x \in X \mid HillClimbing(x) = x^*\}.$$

Multimodal Fitness landscapes and difficulty



The idea :

- if the size of attractive basin of global optimum is "small",
- then, the "time" to find the global optimum is "long"

Optimisation difficulty :

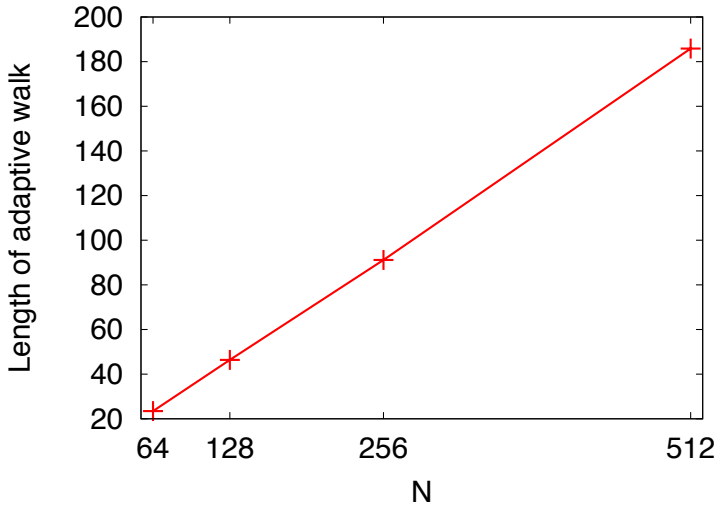
Number and size of attractive basins (Garnier *et al.* [GK02])

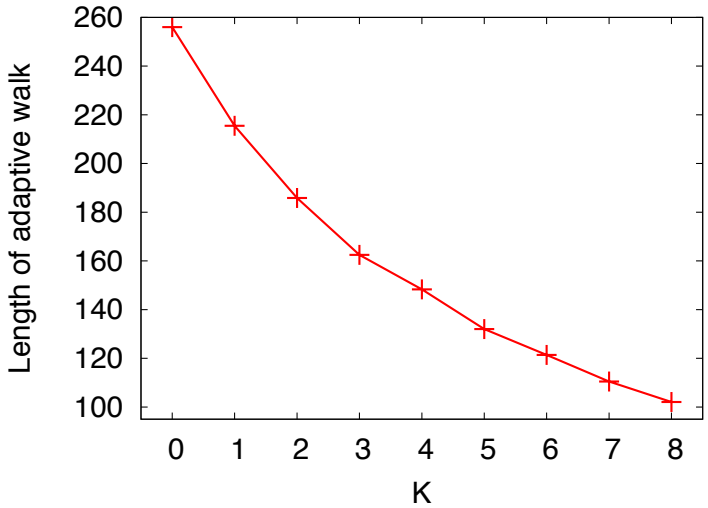
Feature to estimate basin size :

- **Length of adaptive walks**

complexity : sample size $\times \ell \times |\mathcal{N}|$

Multimodal Fitness landscapes and difficulty





ex. nk-landscapes with $n = 512$

The idea :

- if the size of attractive basin of global optimum is "small",
- then, the "time" to find the global optimum is "long"

Optimisation difficulty :

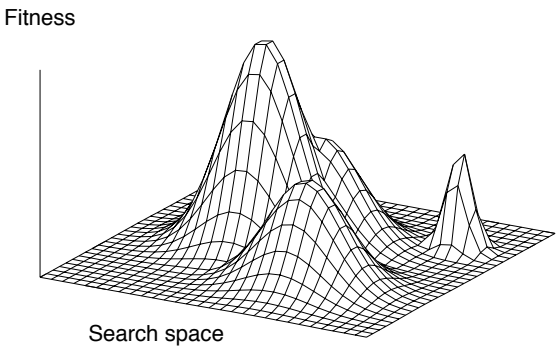
Number and size of attractive basins (Garnier *et al.* [GK02])

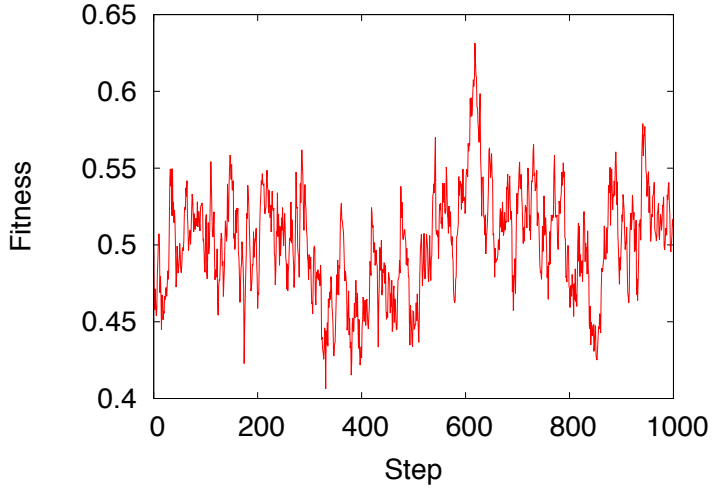
Feature to estimate basin size :

- **Length of adaptive walks**

complexity : sample size $\times \ell \times |\mathcal{N}|$

Random Walk to measure the ruggedness





Random walk :

- $(x_1, x_2, \ldots)$ where $x_{i+1} \in \mathcal{N}(x_i)$ and equiprobability on $\mathcal{N}(x_i)$

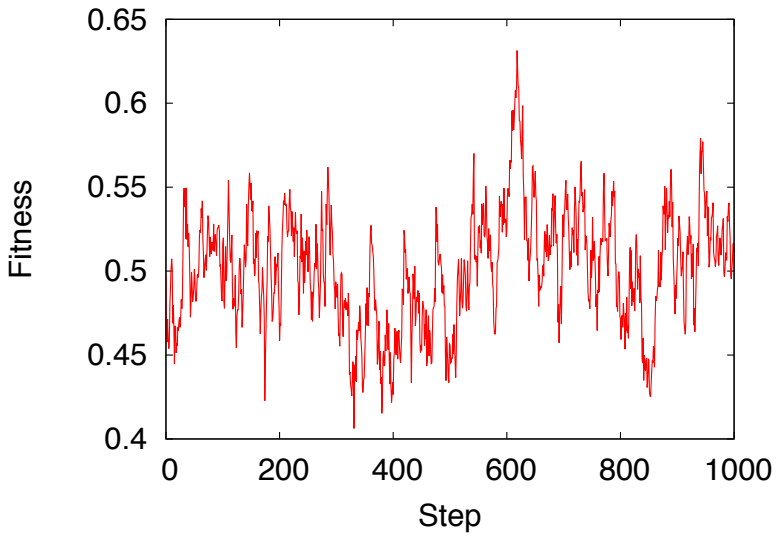
The idea :

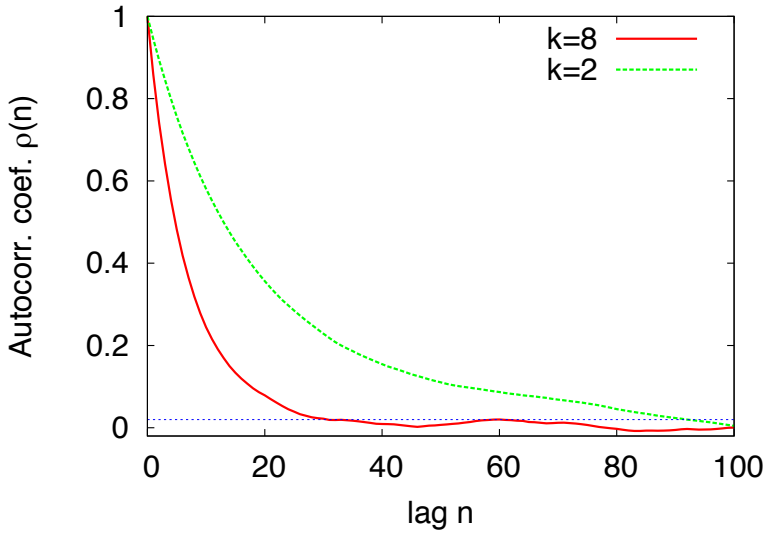
- if the profile of fitness is irregular,
- then, the "information" between neighbors is low.

Feature :

- Study the fitness profile like a signal

Rugged/smooth fitness landscapes





Autocorrelation function of time series of fitnesses along a random walk (Weinberger 90 [Wei90]) :

$$\rho(n) = \frac{\mathrm{E}[(f(x_i) - \bar{f})(f(x_{i+n}) - \bar{f})]}{\mathrm{var}(f(x_i))}$$

Autocorrelation length $\tau = \frac{1}{\rho(1)}$

"How many random steps such that correlation becomes insignificant"

- small τ : rugged landscape
- long τ : smooth landscape

complexity : sample size $\approx 10^3$

Results on rugged fitness landscapes (Stadler 96 [Sta96])

Ruggedness decreases with the size of thoses problems

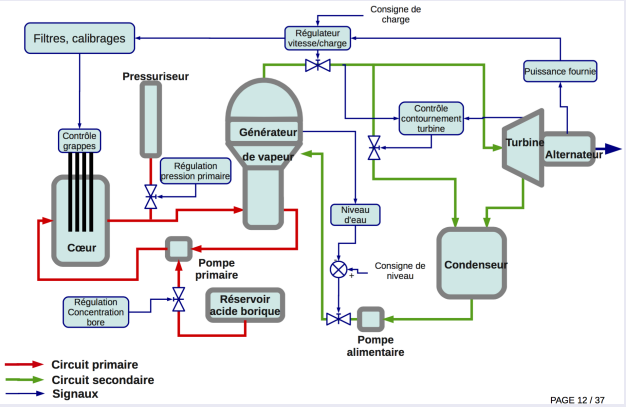
Problem	parameter	$\rho(1)$
symmetric TSP	n number of towns	$1 - \frac{4}{n}$
anti-symmetric TSP	n number of towns	$1 - \frac{4}{n-1}$
Graph Coloring Problem	n number of nodes α number of colors	$1 - \frac{2\alpha}{(\alpha-1)n}$
NK landscapes	N number of proteins K number of epistasis links	$1 - \frac{K+1}{N}$
random max-k-SAT	n number of variables k variables per clause	$1 - \frac{k}{n(1-2^{-k})}$

Tuning using ruggedness features : real-world application

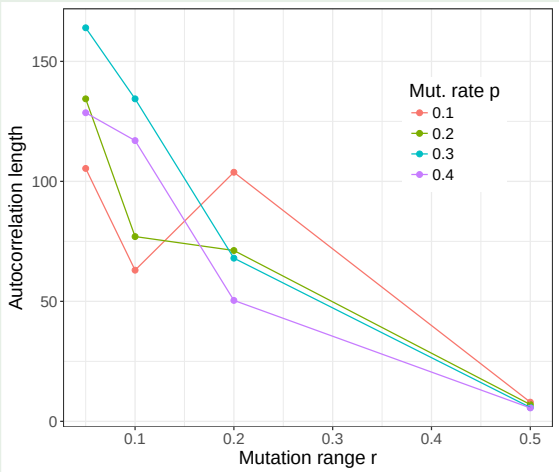
PhD of Mathieu Muniglia, Saclay Nuclear Research Centre (CEA), Paris

New control of nuclear power plant to introduce renewable energies.

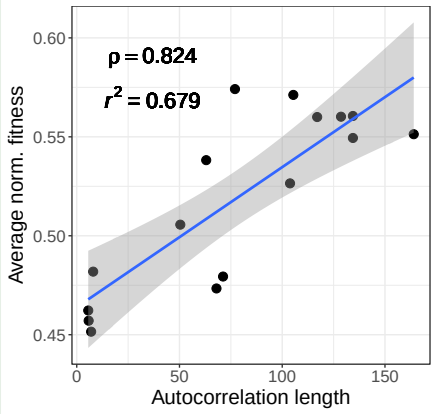
multi-physic simulator
≈ 30min of computation



Parallel M/S EA :
mutation parameters ?



ruggedness depends on parameters



performances is correlated to ruggedness

Offline tuning using fitness landscape

published at evoCOP 2017 :

Towards Landscape-Aware Automatic Algorithm Configuration : Preliminary Experiments on Neutral and Rugged Landscapes.

Arnaud Liefooghe, Bilel Derbel, Sébastien Verel, Hernan Aguirre, and Kiyoshi Tanaka.

See slides presented by Bilel Derbel (thanks to him !)

Joint work

- *Marco Tomassini*, Lausanes University, Switzerland
- *Gabriela Ochoa*, University of Stirling, Scotland
- *Fabio Daolio*, University of Stirling, Scotland

Many thanks !

Key idea : Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

Key idea : Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

$$X \xrightarrow{op} X$$

Variables aggregation for fitness landscape

- At solutions level (small scale) :
  - Stochastic local search operator,
  - Exponential number of solutions,
  - Exponential size of the stochastic matrix of the process (Markov chain)
- Projection on a relevant space :
  - Reduce the size of state space
  - Potentially loose some information
  - Relevant information remains when :

$$p(op(x)) \approx op^{'}(p(x))$$

Key idea : Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

$$\begin{tikzcd} X \arrow[r, "op"] \arrow[d, "p"'] & X \arrow[d, "p"] \\ E \arrow[r, "op'"] & E \end{tikzcd}$$

Variables aggregation for fitness landscape

- At solutions level (small scale) :
  - Stochastic local search operator,
  - Exponential number of solutions,
  - Exponential size of the stochastic matrix of the process (Markov chain)
- Projection on a relevant space :
  - Reduce the size of state space
  - Potentially loose some information
  - Relevant information remains when :

$$p(op(x)) \approx op'(p(x))$$

Key idea : Complex system tools

Complex network

Bring the tools of *complex networks* analysis to the study the structure of combinatorial fitness landscapes

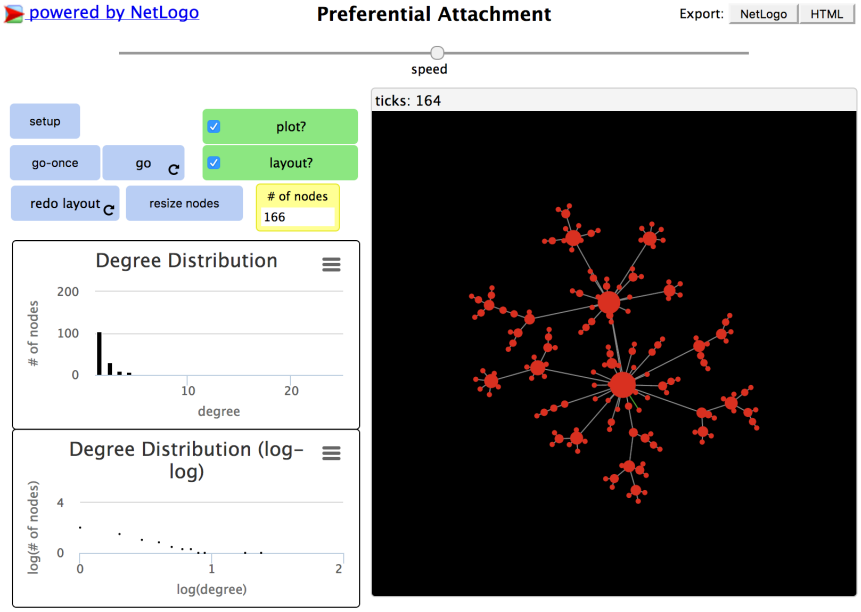
Methodology

- **Design a network** that represents the landscape
  - Vertices : local optima
  - Edges : a notion of adjacency between local optima
- **Extract features** :
  "complex" network analysis
- **Use the network features** :
  search algorithm design, difficulty, etc.

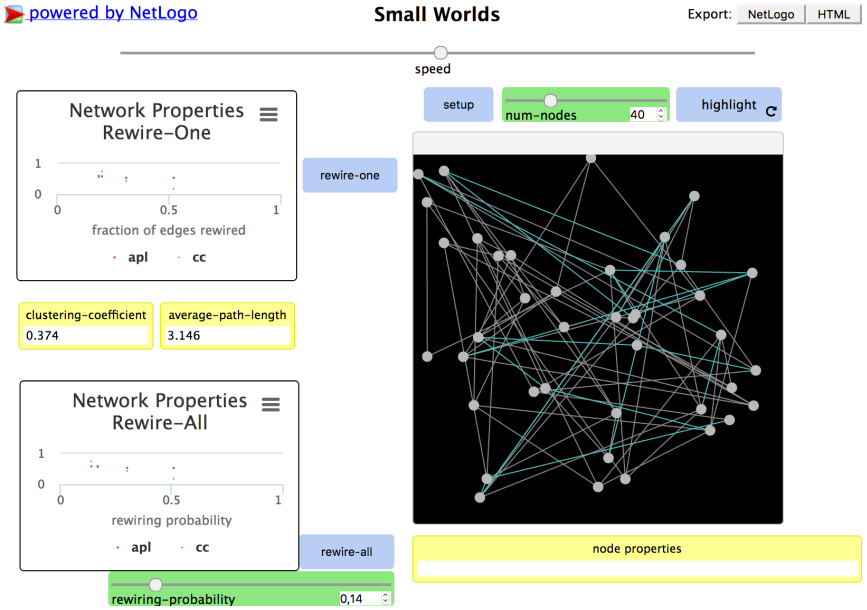
J. P. K. Doye, The network topology of a potential energy landscape : a static scale-free network., *Phys. Rev. Lett.*, 88 :238701, 2002. [Doy02]

Complex networks

Scale free network
(Watts and Strogatz, 1998 [WS98])



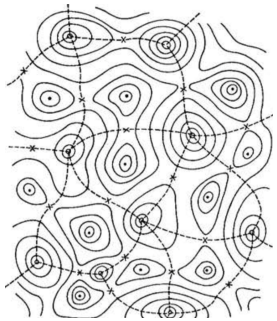
Small world network
(Barabasi and Albert, 1999 [BA99])



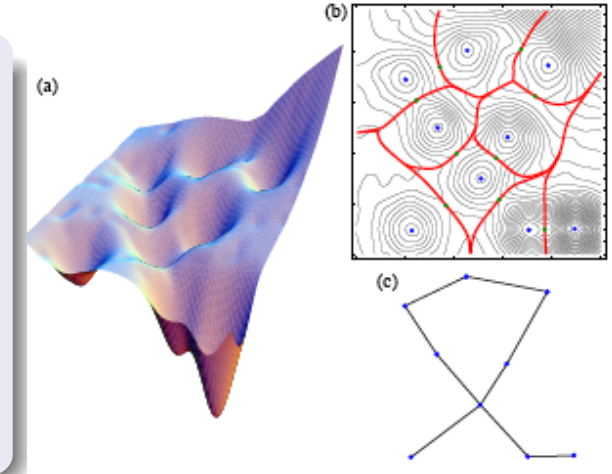
Energy surface and inherent networks

Inherent network

- **Nodes** : energy minima
- **Edges** : two nodes are connected if the energy barrier separating them is sufficiently low (transition state)



(a) Energy surface
(b) Contours plot : partition of states space into basins of attraction
(c) Landscape as a network



F. H Stillinger, T. A Weber. Packing structures and transitions in liquids and solids. *Science*, 225.4666 , p. 983-9, 1984.[SW84]
J. P. K. Doye, The network topology of a potential energy landscape : a static scale-free network., *Phys. Rev. Lett.*, 88 :238701, 2002.[Doy02]

Local Optima Network

Definition : Local Optima Network (LON)

Orienter weighted graph (V, E, w)

- Notes V : set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges E : notion of connectivity between local optima

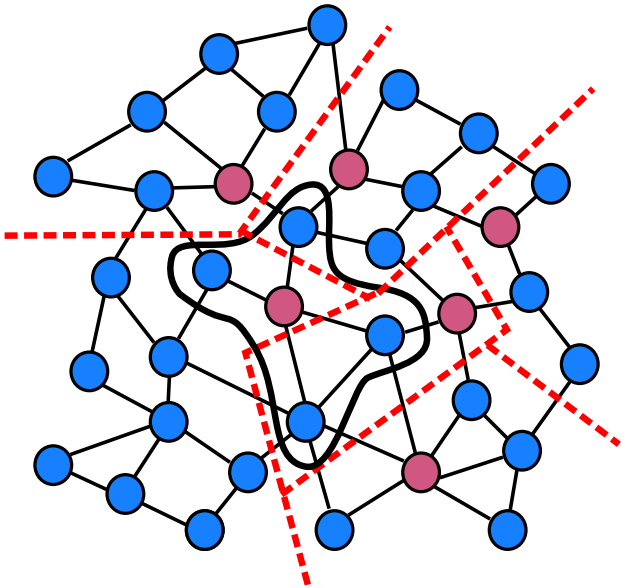
Escape edges

Edge e_{ij} between LO_i and LO_j
if $\exists x : distance(LO_i, x) \leqslant D$ and $x \in b_j$.

Weights

$w_{ij} = \sharp\{x \in X \mid d(LO_i, x) \leqslant D,\ x \in b_j\}$
can be normalized by the number of solutions at

distance D



Local Optima Network

Definition : Local Optima Network (LON)

Orienter weighted graph (V, E, w)

- Notes V : set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges E : notion of connectivity between local optima

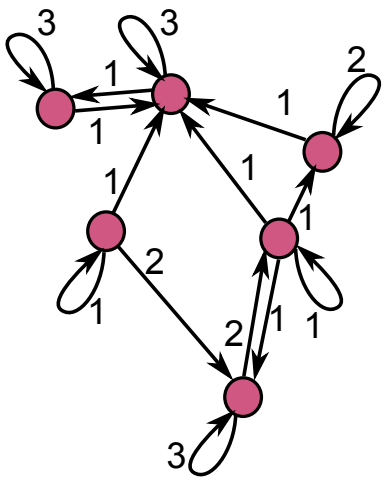
Escape edges

Edge e_{ij} between LO_i and LO_j
if $\exists x : distance(LO_i, x) \leqslant D$ and $x \in b_j$.

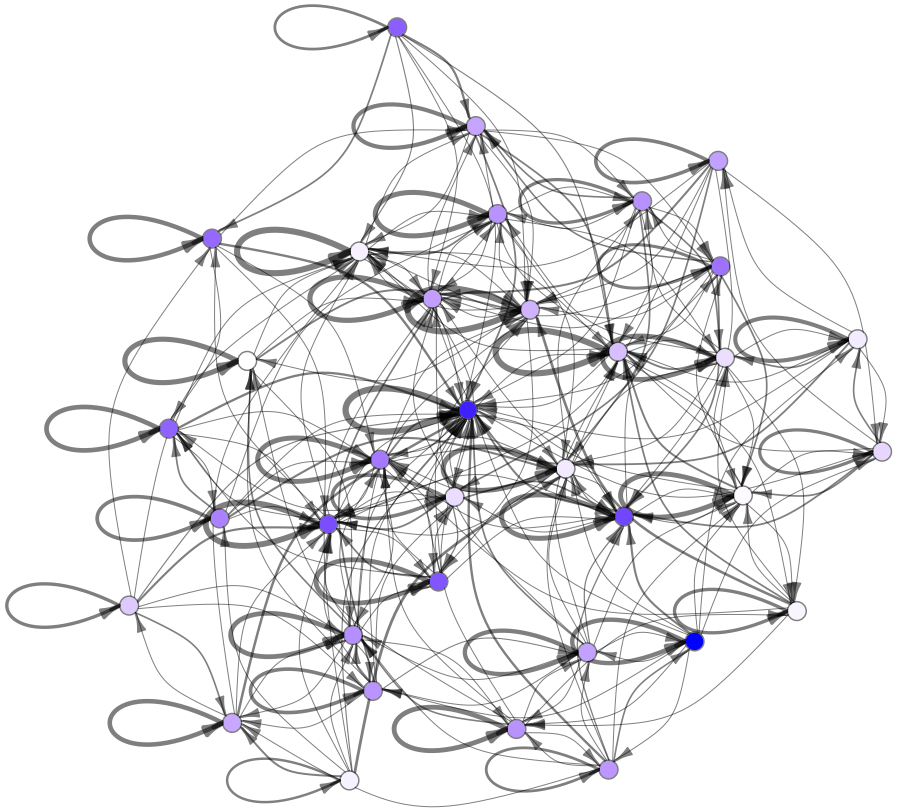
Weights

$w_{ij} = \sharp\{x \in X \mid d(LO_i, x) \leqslant D,\ x \in b_j\}$
can be normalized by the number of solutions at

distance D



Features of local optima network



- nv : $\sharp$vertices
- lv : avg path length
  $d_{ij} = 1/w_{ij}$
- lo : path length to best
- fnn : fitness corr.
  $(f(x), f(y))$ with $(x, y) \in E$
- wii : self loops
- wcc : weighted clust. coef.
- $zout$: out degree
- $y2$: disparity
- knn : degree corr.
  $(deg(x), deg(y))$ with $(x, y) \in E$

Benchmark : NK-landscapes

[Kauffman 1993] [Kau93]

$$x \in \{0,1\}^n \qquad f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x_j, x_{i_1}, \ldots, x_{i_k})$$

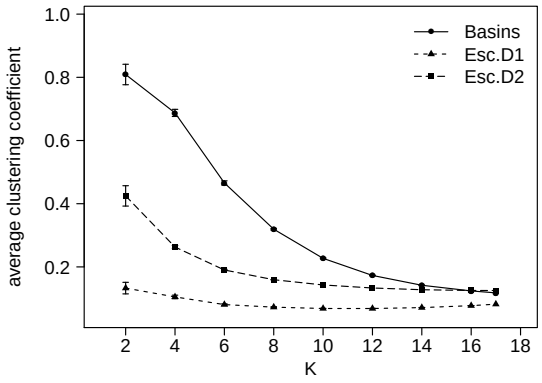
Two parameters :

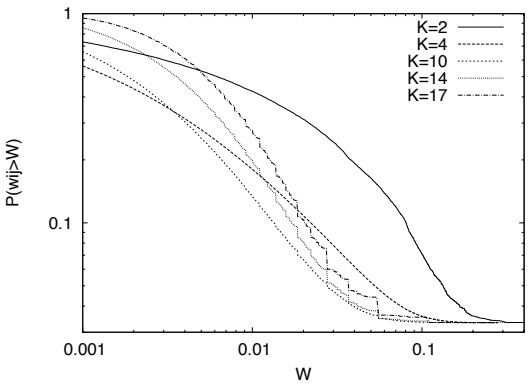
- Problem size n
- Non-linearity k $<$ n
  (multi-modality, epistatic interactions)
  - $k = 0$: linear problem, one single maxima
  - $k = n - 1$: random problem, number of local optima $\frac{2^N}{N+1}$

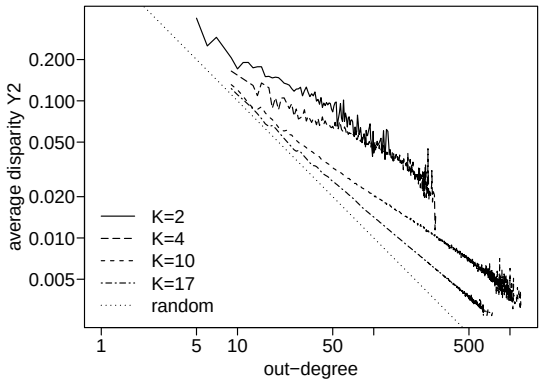
remarks : "same" results with QAP, flow shop.

Structure of Local Optima Network

- NK-landscapes (small instances) :
  Most of the features are correlated with K
  relevance of LON definition







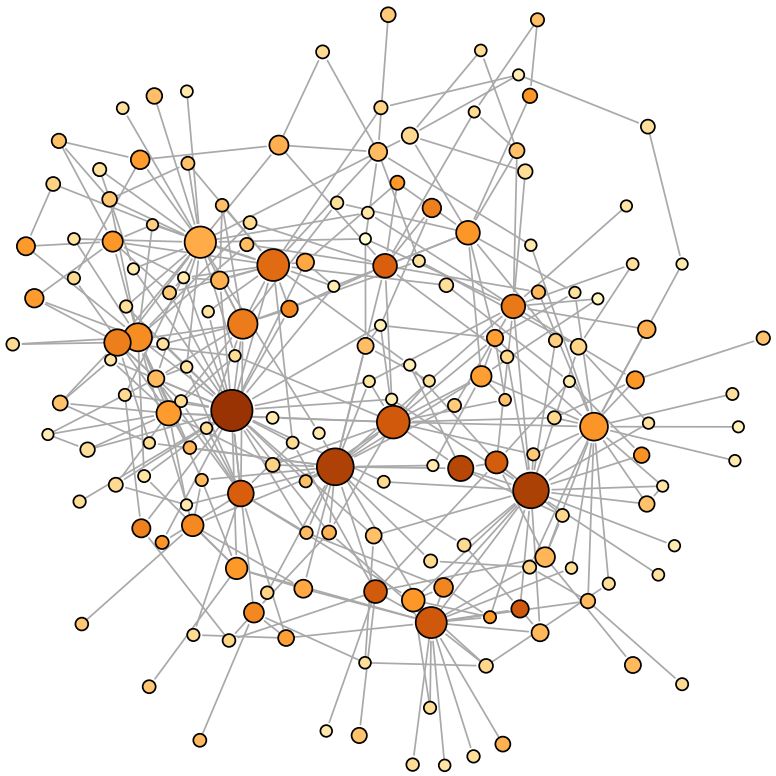
- LON is **not a random** network (NK, QAP, FSSP) :
  Highly clustered network,
  Distribution of weights and degrees have long tail, etc.

LON to compare of problem difficulty

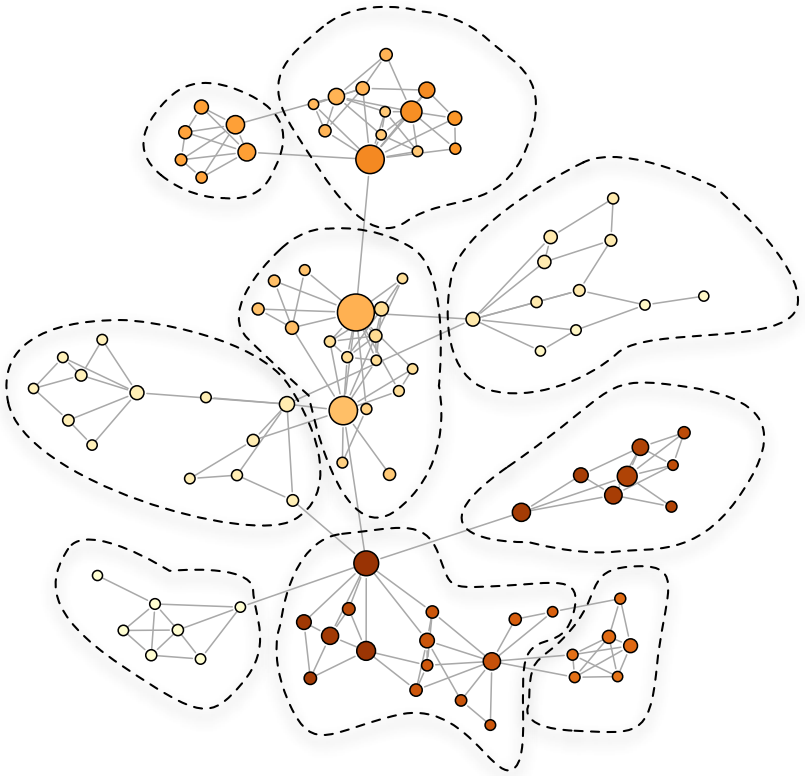
Local Optima Network for Quadratic Assignment Problem (QAP) [DTVO11]

→ Community detection in LON for

Random instance



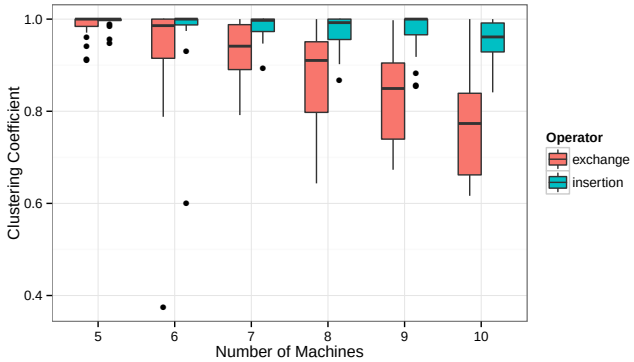
Real-like instance

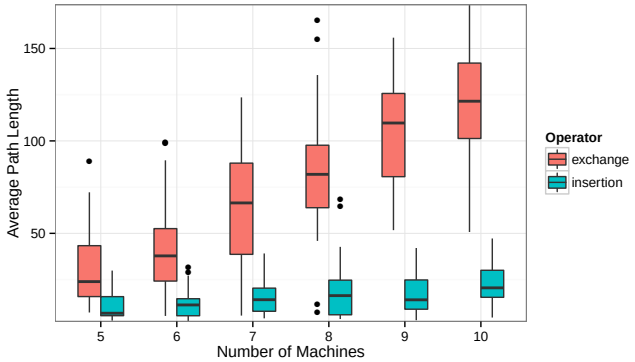


Structure of the LON related to **problem difficulty**

LON to compare algorithm components

Comparaison of **operators** for Flow Shop Scheduling Problem





Comparaison of **pivot rule** in hill-climbing for NK-landscapes

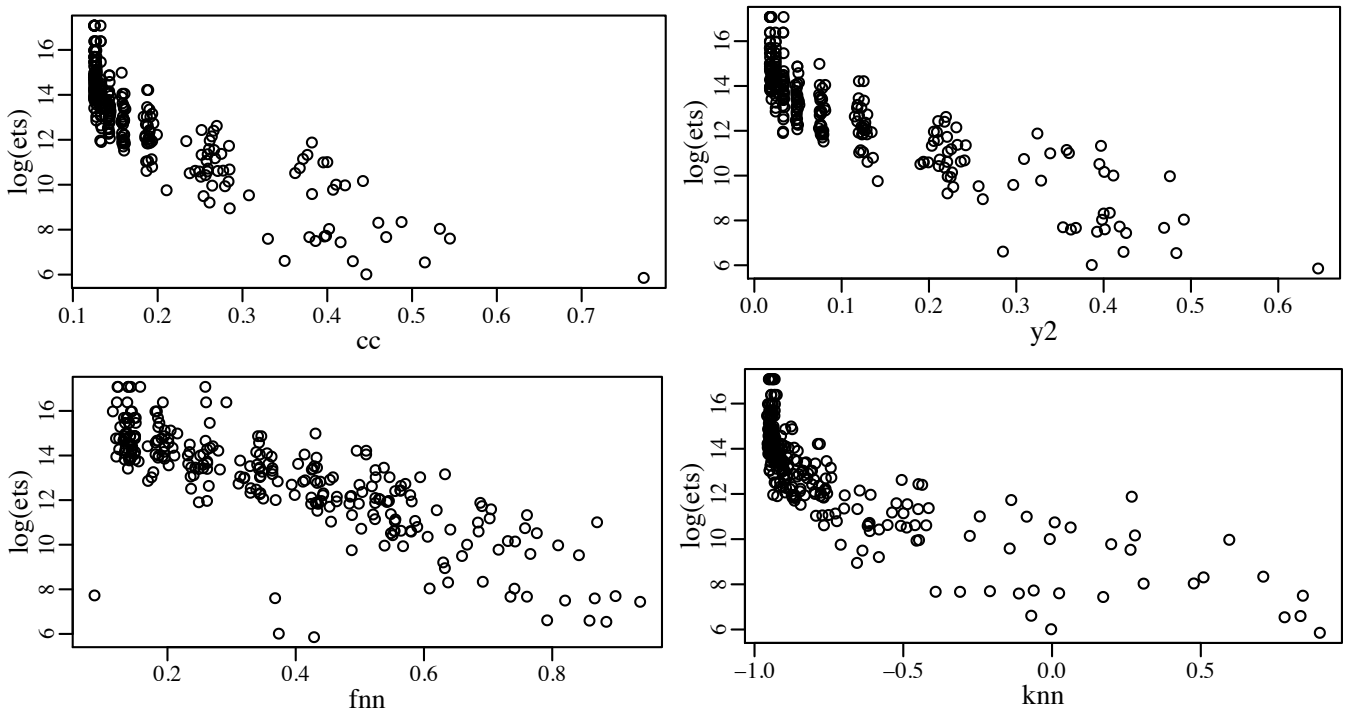
K	$\bar{n}_e/\bar{n}_v^2$		$\bar{Y}$		$\bar{d}$		$\bar{d}_{best}$	
	b-LON	f-LON	b-LON	f-LON	b-LON	f-LON	b-LON	f-LON
2	0.81	0.96	0.326	0.110	56	39	16	12
4	0.60	0.92	0.137	0.033	126	127	35	32
6	0.32	0.79	0.084	0.016	170	215	60	70
8	0.17	0.65	0.062	0.011	194	282	83	118
10	0.09	0.53	0.050	0.009	206	340	112	183

LON features *vs.* performance : simple correlation

Algorithm : Iterated Local Search on NK-landscapes with $N = 18$

Performance : $ert = \mathbb{E}(T_s) + \left(\frac{1-p_s}{p_s}\right) T_{max}$

n_v	$\bar{d}_{best}$	$\bar{d}$	fnn	w_{ii}	$\bar{C}^w$	zout	$\bar{Y}$	knn
0.885	0.915	0.006	−0.830	−0.883	−0.875	0.885	−0.883	−0.850



LON features *vs.* performance : multi-linear regression

1. Multiple **linear** regression on all possible predictors :

$$\log(ert) = \beta_0 + \beta_1 k + \beta_2 \log(nv) + \beta_2 lo + \cdots + \beta_{10} knn + \varepsilon$$

2. Step-wise **backward elimination** of each predictor in turn.

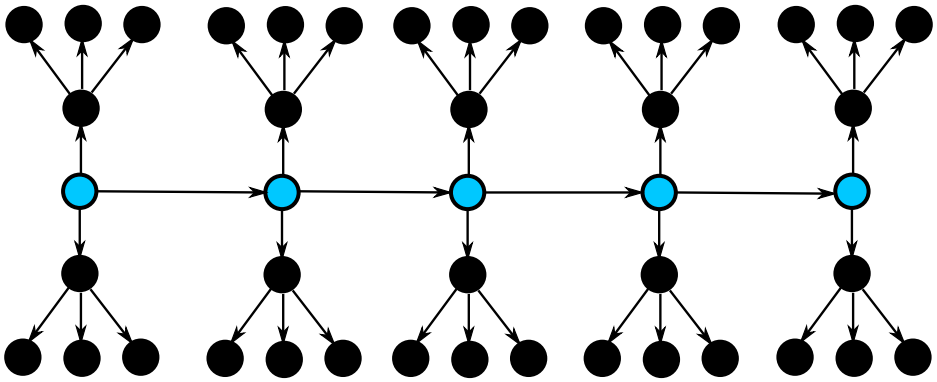
Predictor	$\widehat{\beta}_i$	Std. Error	p-value
(Intercept)	10.3838	0.58512	$9.24 \cdot 10^{-47}$
lo	0.0439	0.00434	$1.67 \cdot 10^{-20}$
zout	−0.0306	0.00831	$2.81 \cdot 10^{-04}$
y2	−7.2831	1.63038	$1.18 \cdot 10^{-05}$
knn	−0.7457	0.40501	$6.67 \cdot 10^{-02}$

Multiple R-squared : 0.8494, Adjusted R-squared : 0.8471.

Sampling methodology for large size instances

From the sampling of large-size complex network :

- Random walk on the network
- Breadth-First-Search

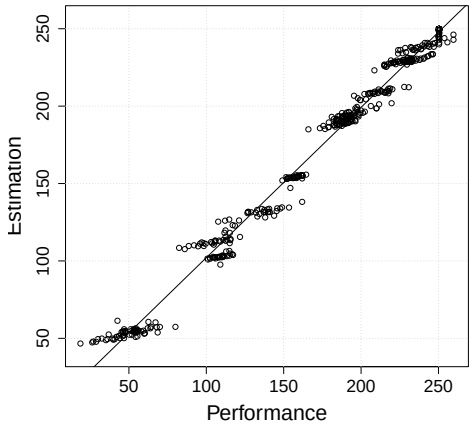


Procedure LONSampling(d, m, l)
$x_0 \leftarrow hc(x)$ with x random solution
for $t \leftarrow 0, \ldots l-1$ **do**
  Snowball(d, m, x_t)
  $x_{t+1} \leftarrow$ RandomWalkStep(x_t)
end for

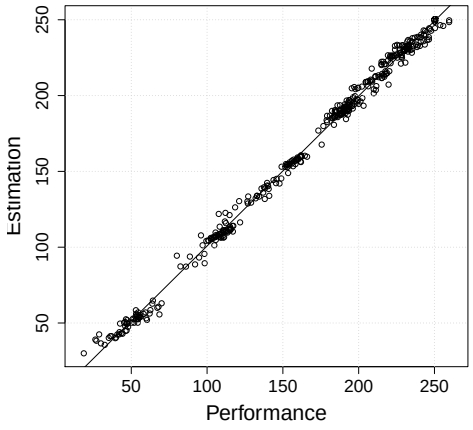
Performance prediction based on estimated features

- Optimization scenario using off-the-shelf metaheuristics :
  TS, SA, EA, ILS on 450 instances for NK and QAP.
- Performance measures :
  average fitness / average rank
- Model of regression :
  linear model / random forest
- Set of features :
  - *basic* : 1st autocorr. coeff. of fitness (rw of length 10^3),
    Avg. fitness of local optima (10^3 hc),
    Avg. length to reach local optima (10^3 hc).
  - *lon* : see previous,
  - *all* : *basic* and *lon* features
- Quality measure of regression :
  R^2 on cross-validation (repeated random sub-sampling)

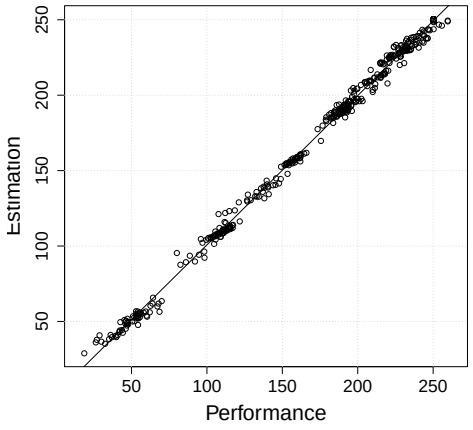
Scatter plots of the observed-estimated performance



basic, $R^2 = 0.9327$



lon, $R^2 = 0.9601$



all, $R^2 = 0.9643$

Very good prediction using LON features

Algorithm selection : portfolio scenario

- Portfolio of 4 metaheuristics : TS, SA, EA, ILS
- Classification task : selection of one of the best metaheuristic
- Models : logit, random forest, svm
- Quality of classification :
  error rate (algo. is not one of the best) on cross-validation.

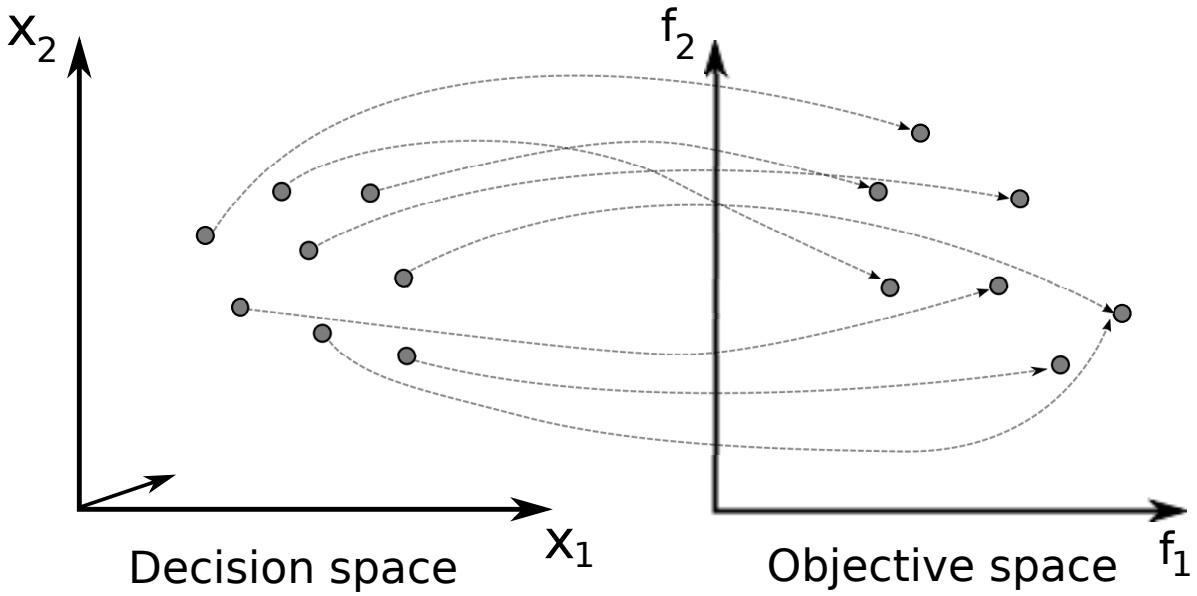
		Avg. error rate				
Probl.	Feat.	logit	rf	svm	cst	rnd
NK	basic	0.0379	0.0278	**0.0158**	0.4711	0.6749
	lon	**0.0203**	**0.0249**	**0.0168**		
	all	0.0244	**0.0269**	**0.0165**		
QAP	basic	**0.0142**	0.0107	0.0771	0.4222	0.6706
	lon	**0.0156**	**0.0086**	**0.0456**		
	all	**0.0161**	0.0106	**0.0431**		

Multiobjective optimization

Multiobjective optimization problem

- Search space : *decision space* X
- Objective function : *objective space* $f : X \rightarrow \mathbb{R}^m$

component f_i : objective, criterion.

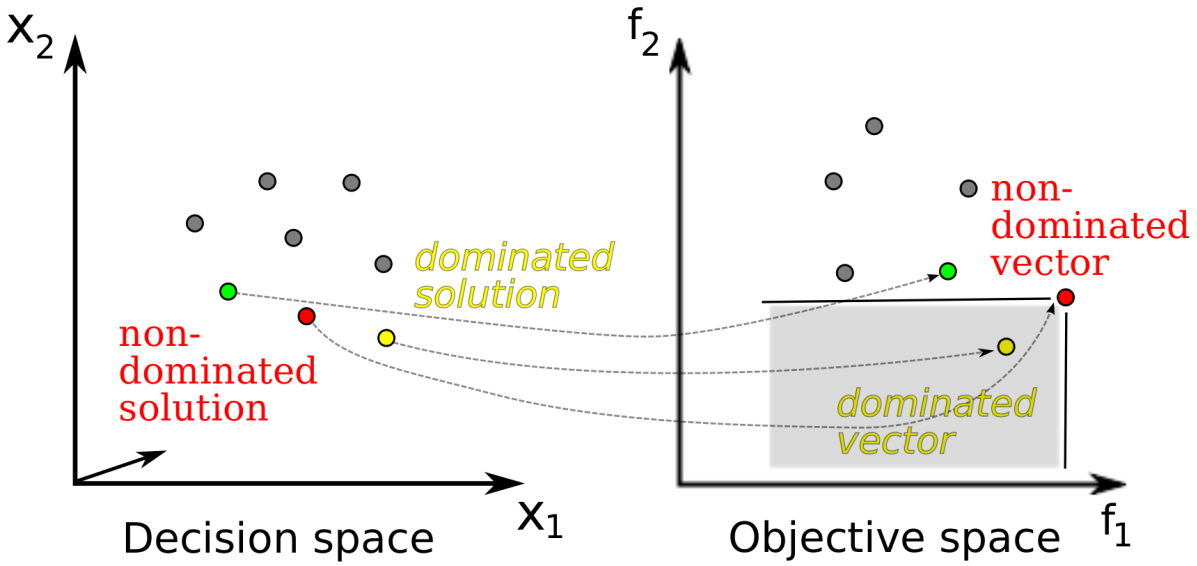


Definition

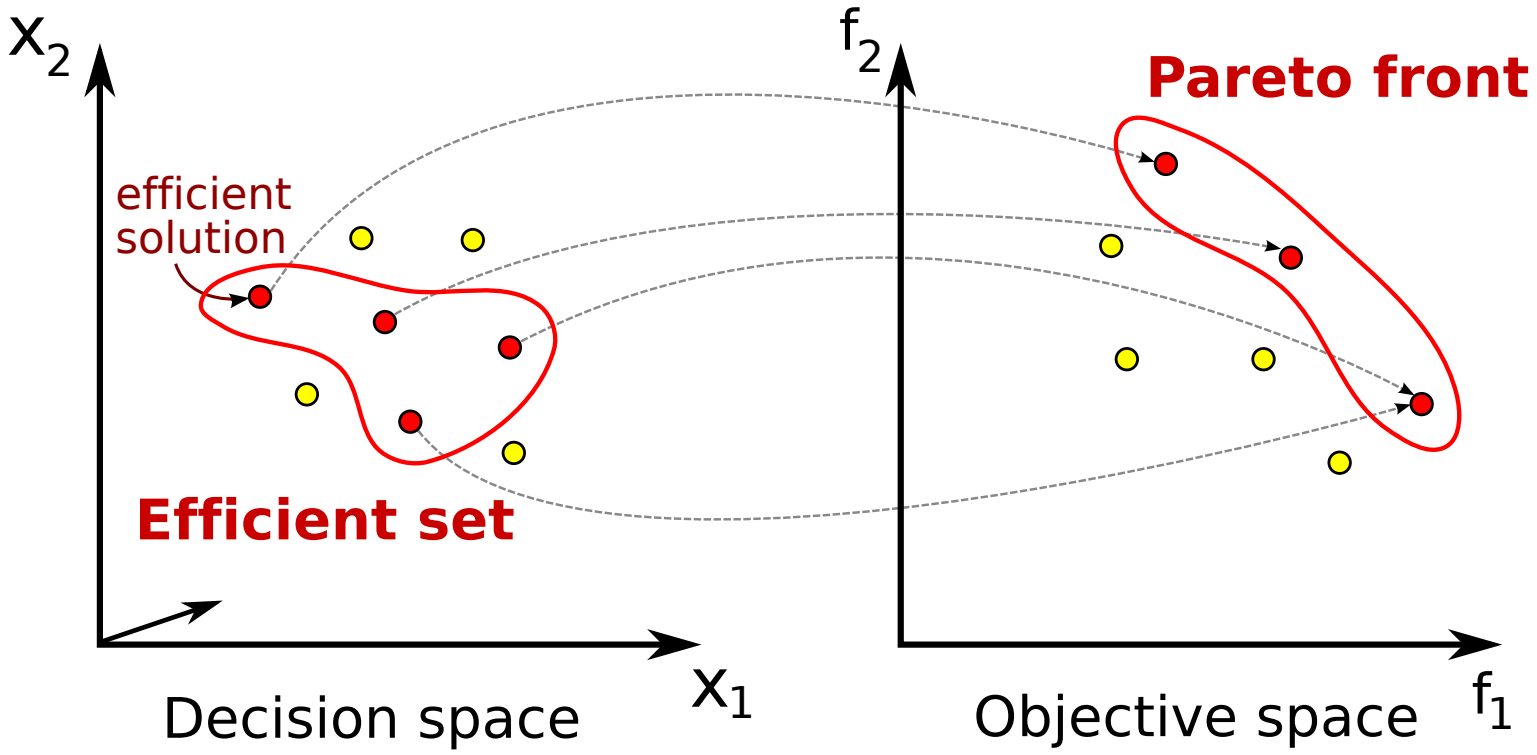
Pareto dominance relation

A solution $x \in X$ dominates a solution $x' \in X$ ($x' \prec x$) iff

- $\forall i \in \{1, 2, \ldots, M\}$, $f_i(x') \leqslant f_i(x)$
- $\exists j \in \{1, 2, \ldots, M\}$ such that $f_j(x') < f_j(x)$



Pareto set, Pareto front



Goal of multi-objective optimization

Find the whole **Pareto Optimal Set**,
or a good approximation of the Pareto Optimal Set

Fitness landscape for Multi-Objective optimization

Motivation

- Multi-objective optimization problems are hard :
  underlying single-objective functions, joined functions, etc.
- Learning the problem structure :
  to understand problem difficulty
  to improve algorithm design
  to explain and predict algorithm performance
- Questions :
  - What are the relevant problem features ?

Joint work

- *Arnaud Liefooghe*, University of Lille/inria, France
- *Fabio Daolio*, University of Stirling, UK
- *Hernán Aguirre*, Shinshu University, Japan
- *Kiyoshi Tanaka*, Shinshu University, Japan



MODŌ
INTERNATIONAL ASSOCIATED LABORATORY
SHINSHU univ + univ LILLE

Slides are from the team, many thanks !

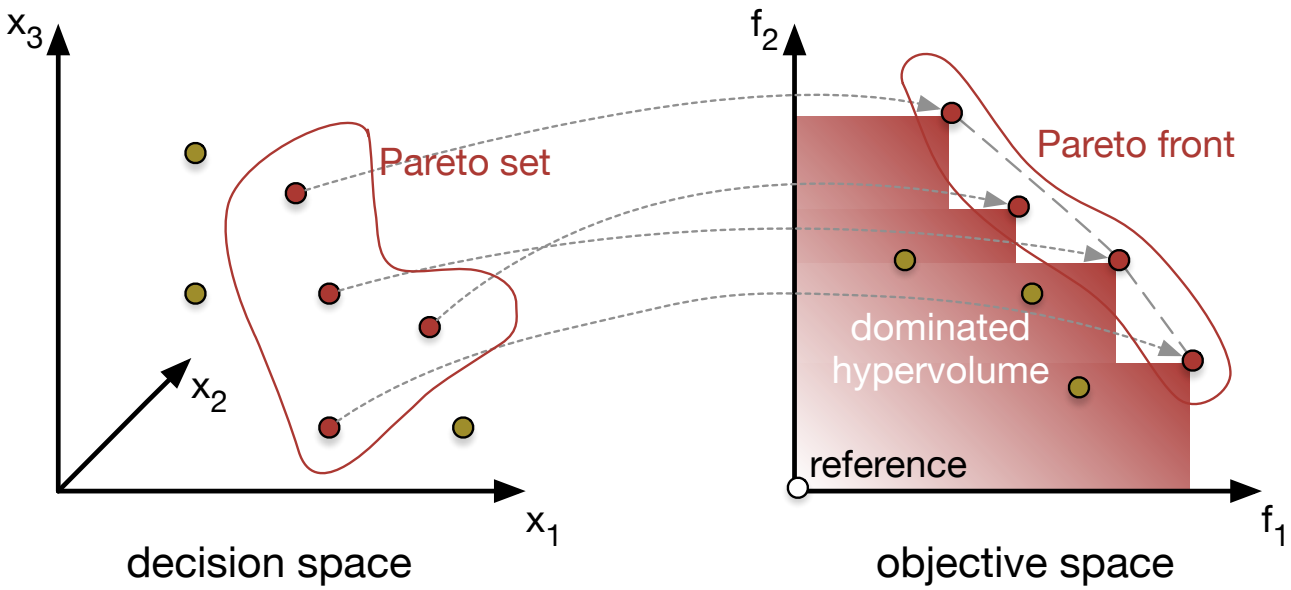
What is a good Pareto set approximation ?

Rule of thumb

- closeness to the (exact) Pareto front
- well-distributed solutions in the objective space
- the more, the better ?

Quality indicators

- scalar value that reflects an approximation set quality
- IGD, EPS, R-metrics, HV . . . (all have limitations and biases)



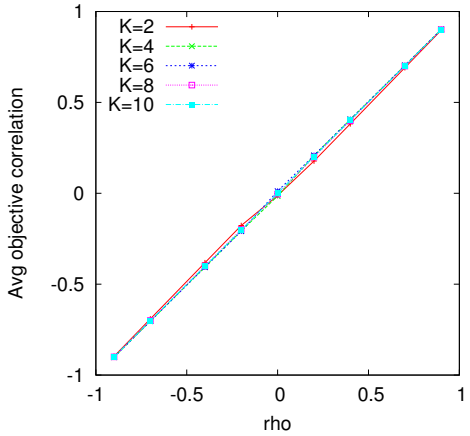
ρMNK-landscapes [Kauffman 1993 ; Aguirre & Tanaka 2007 ; Verel et al. 2010]

general-purpose family of multi-modal pseudo-boolean optimization functions

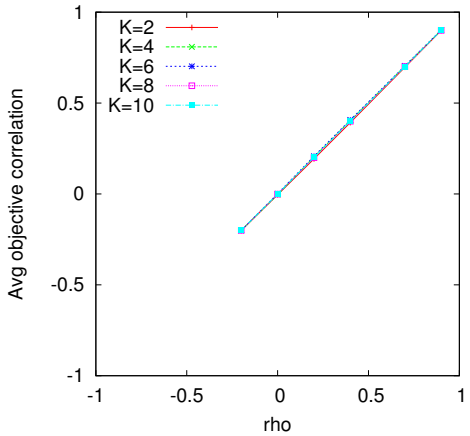
superposition of n Walsh functions of order k+1

$$\begin{array}{lll} \max & f_i(x) = \frac{1}{n}\sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) & , \quad i \in \{1, \ldots, m\} \\ \text{s.t.} & x_j \in \{0, 1\} & , \quad j \in \{1, \ldots, n\} \end{array}$$

m = 2



m = 5

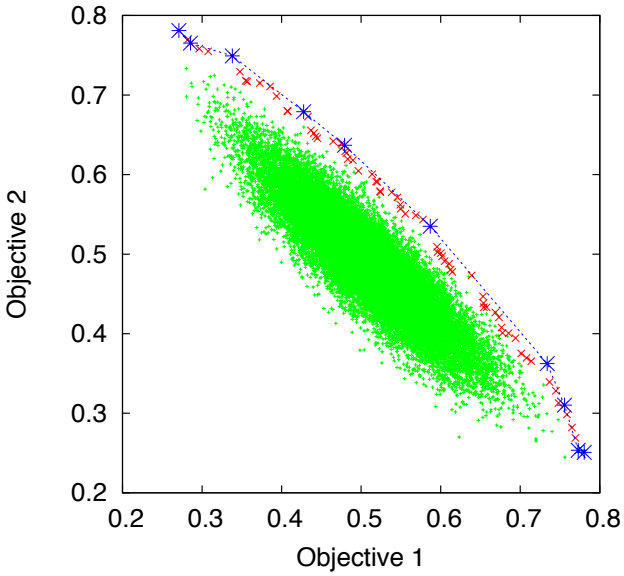


Benchmark parameters :

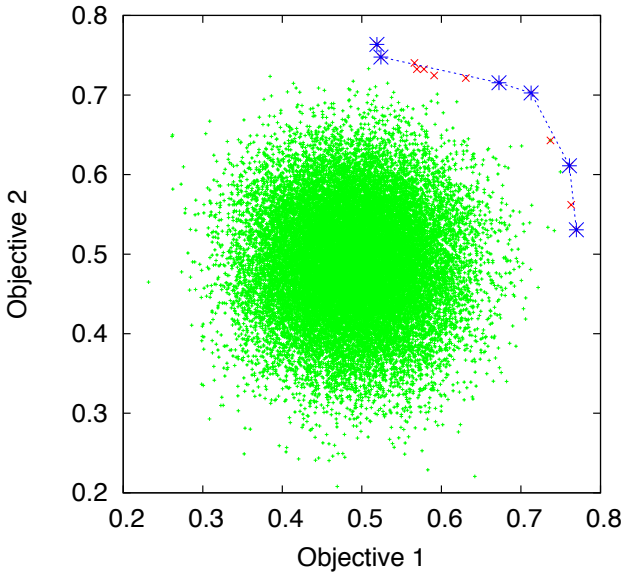
- problem size n (decision space dimension)
- problem non-linearity $k <$ n (multi-modality, epistatic interactions)
- number of objective functions m (objective space dimension)
- objective correlation $\rho > -\frac{1}{m-1}$

http://mocobench.sf.net

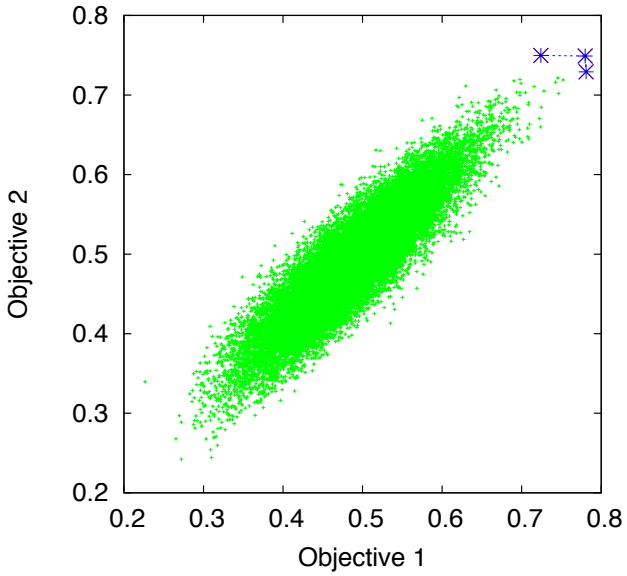
Some intuitions on objective correlation ρ



conflicting objectives
$\rho = -0.9$



independent objectives
$\rho = 0.0$



correlated objectives
$\rho = 0.9$

$m = 2 \qquad n = 18 \qquad k = 4$

EMO algorithm classes

Scalarizing approaches

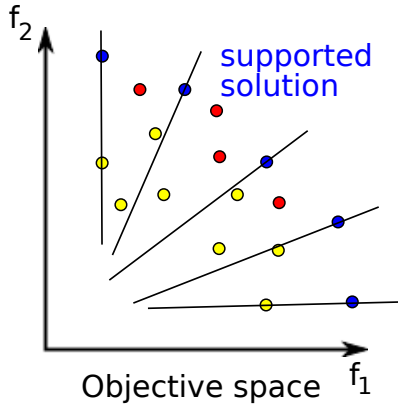
- multiple aggregations of the objectives (e.g. weighted-sum)
- beware of **unsupported** solutions
- MOSA, MOTS, TPLS, MOEA/D ...

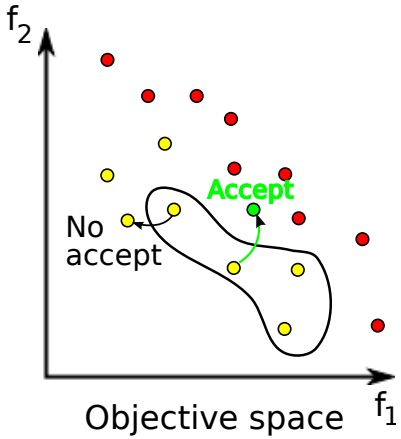
Dominance-based approaches

- search process guided by a dominance relation
- NSGA-II, SPEA2, PAES, PLS, SEMO, AεSεH ...

Indicator-based approaches

- search process guided by a quality indicator
- IBEA, IBMOLS, SMS-EMOA, HypE ...





Two prototypical dominance-based EMO algorithms

local search
multi-objective hill-climber
PLS
[Paquete et al. 2004]

global search
multi-objective $(1+1)-$EA
G-SEMO
[Laumanns et al. 2004]

$A \leftarrow \{x_0\}$
repeat
  select $x \in A$ at random
  for all x' s.t. $||x - x'||_1 = 1$ **do**
    $A \leftarrow$ non-dominated
      solutions from $A \cup \{x'\}$
  end for
until stop

$A \leftarrow \{x_0\}$
repeat
  select $x \in A$ at random
  $x' \leftarrow x$
  flip each bit x'_i with a rate $\frac{1}{n}$
  $A \leftarrow$ non-dominated
    solutions from $A \cup \{x'\}$
until stop

Benchmark parameters

Parameters from ρMNK-landscapes

n problem size
(solution space dimension)

k problem non-linearity
(number of epistatic interactions)

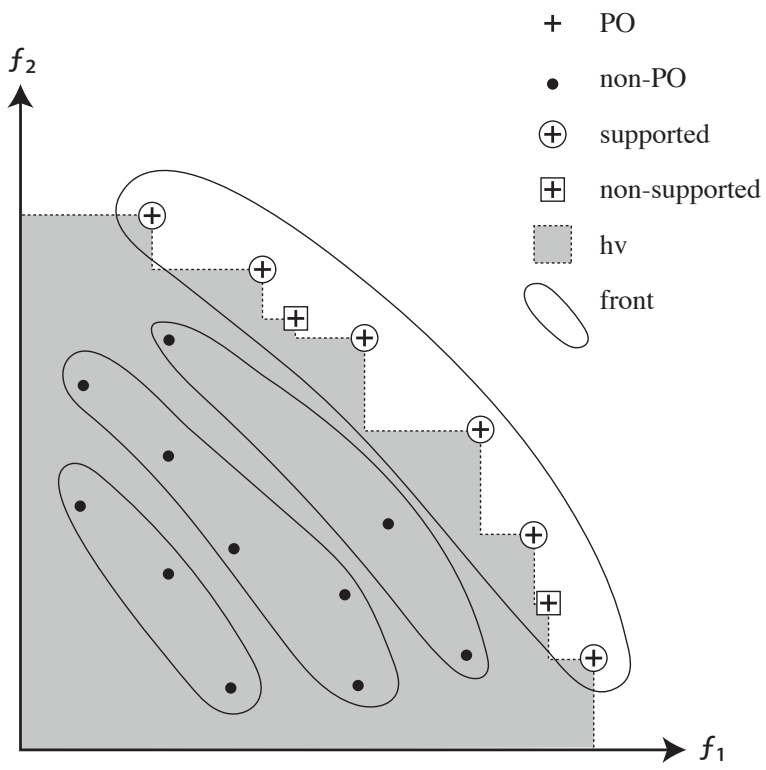
m number of objective functions
(objective space dimension)

ρ objective correlation
(correlation between the objective function values)

$$\begin{array}{lll} \max & f_i(x) = \frac{1}{n}\sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) & , \quad i \in \{1, \ldots, m\} \\ \text{s.t.} & x_j \in \{0, 1\} & , \quad j \in \{1, \ldots, n\} \end{array}$$

Global features from full enumeration (1)

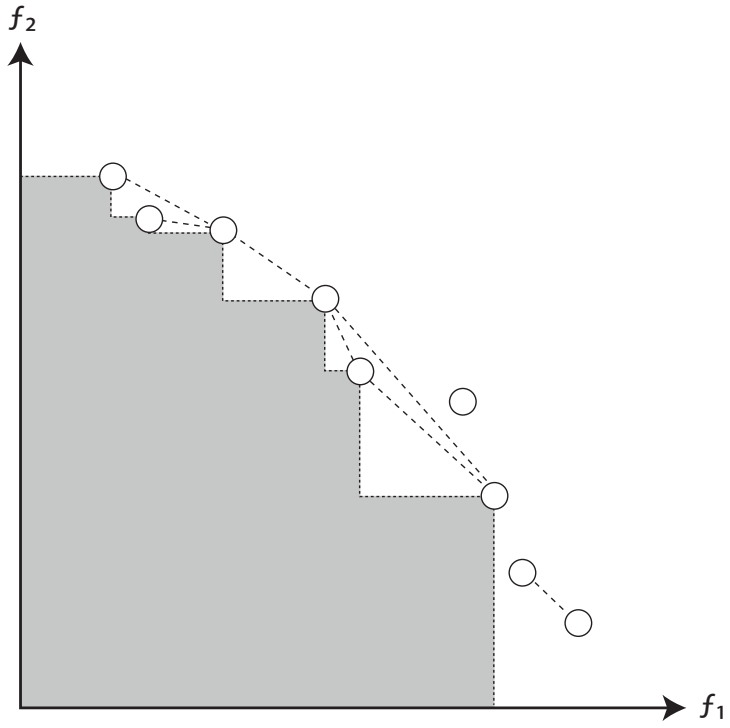
Features from the Pareto set/solution space



#po	Pareto optimal (PO) sol.
#supp	supported PO solutions
hv	PF's hypervolume
#fronts	non-dominated fronts
front_ent	entropy of front's size distribution

Global features from full enumeration (2)

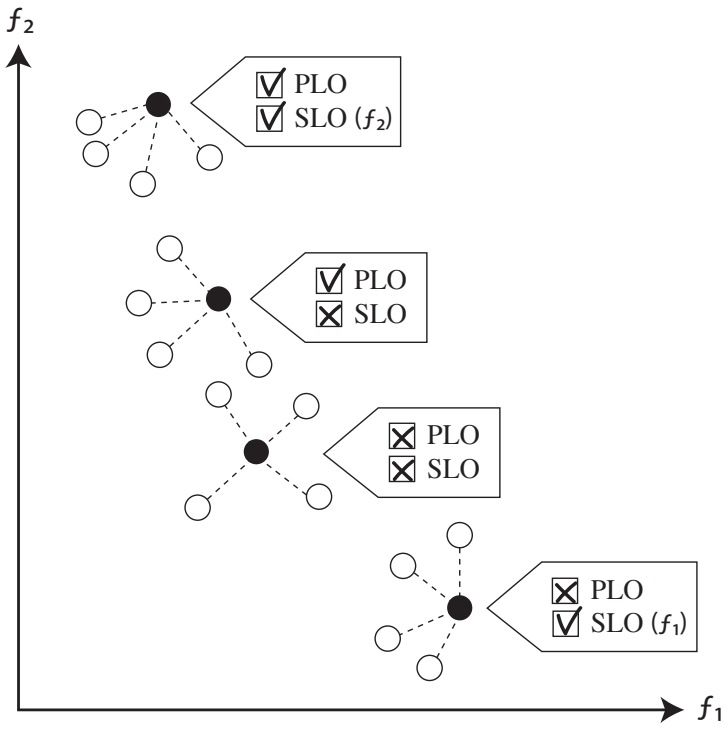
Features from the Pareto set/graph



podist_avg	avg Hamming distance
podist_max	max distance (diameter)
fdc	fitness-distance correlation
#cc	connected components
#sing	singletons
#lcc	largest connected comp.
lcc_dist	avg distance in LCC
lcc_hv	LCC's hypervolume

Global features from full enumeration (3)

Local optimality

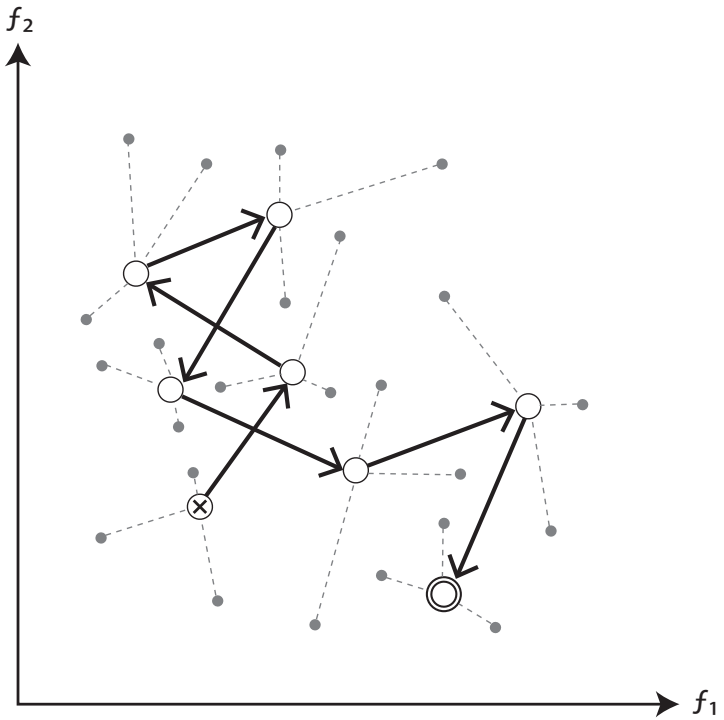


`#plo` Pareto local optimal (PLO) solutions

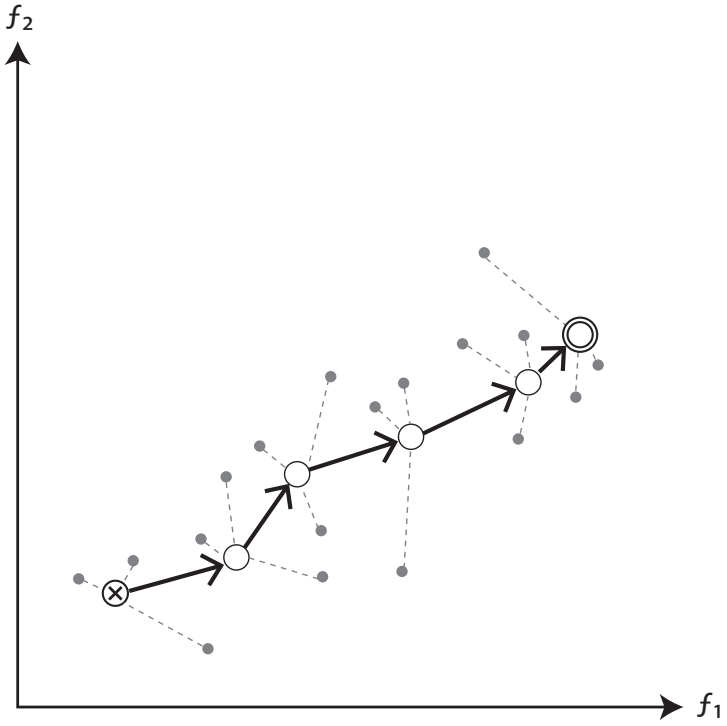
`#slo_avg` single-objective local optima (SLO) per objective (avg)

Local features from sampling (1)

Multi-objective random/adaptive walk



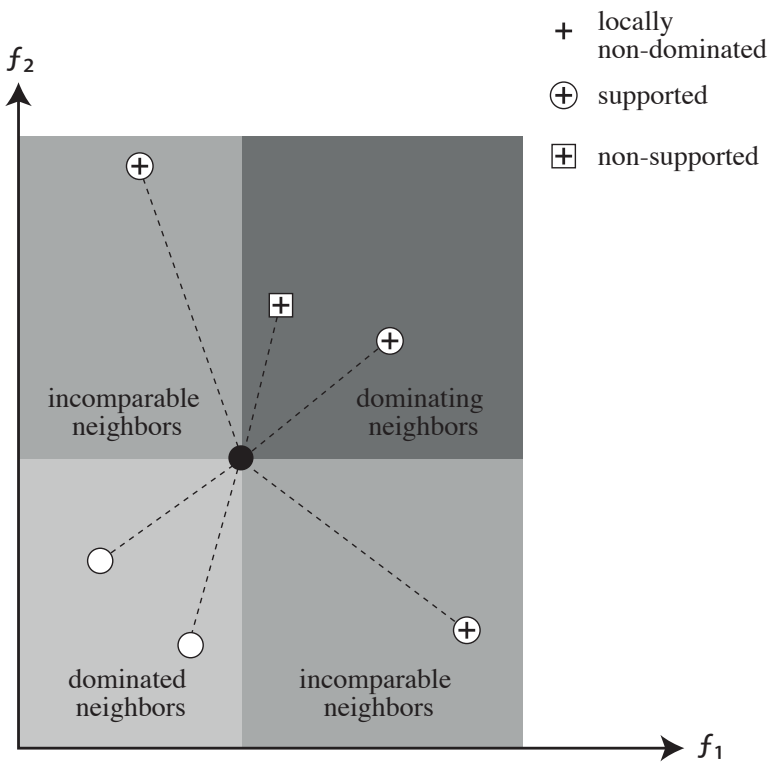
random walk sampling (*rws*)



adaptive walk sampling (*aws*)

Local features from sampling (2)

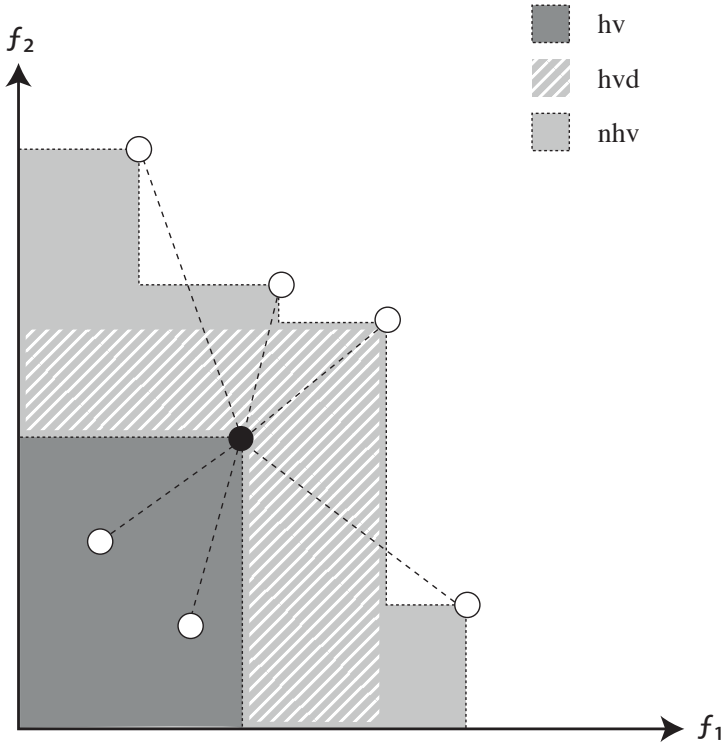
Dominance-based metrics



- locally non-dominated solutions in the neighborhood
- supported locally non-dominated solutions in the neighborhood
- neighbors dominated by the current solution
- neighbors dominating the current solution
- neighbors incomparable to the current solution
- average length of aws

Local features from sampling (3)

Hypervolume-based metrics



- (single) solution's hypervolume
- (single) solution's hypervolume difference
- neighborhood's hypervolume

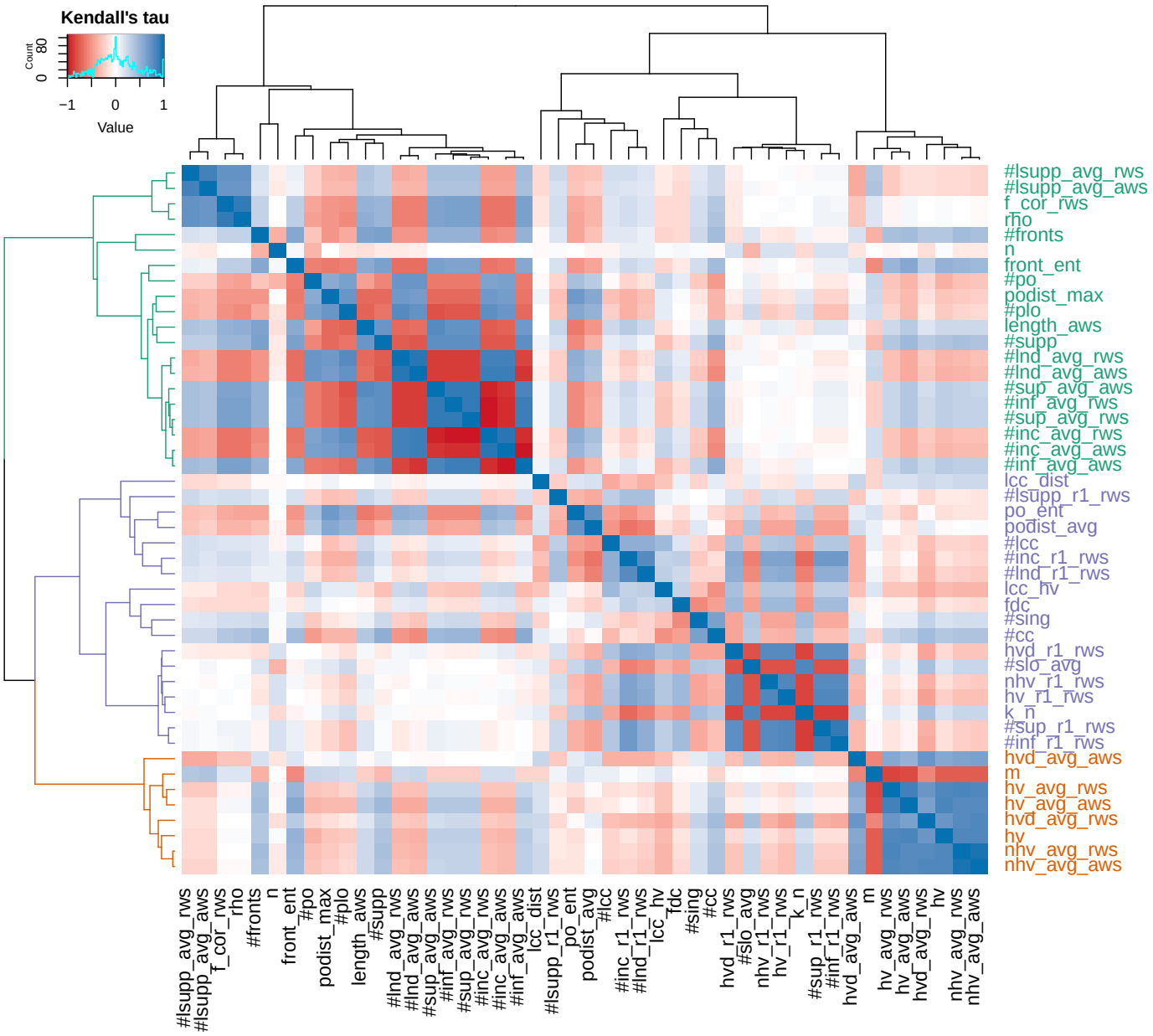
Summary of problem features (1)

	BENCHMARK **parameters** (4)	
n	number of (binary) variables	
k_n	proportional number of variable interactions (epistatic links) : k/n	
m	number of objectives	
ρ	correlation between the objective values	
	GLOBAL FEATURES FROM **full enumeration** (16)	
#po	proportion of Pareto optimal (PO) solutions	*knowles2003*
#supp	proportion of supported solutions in the Pareto set	*knowles2003*
hv	hypervolume-value of the (exact) Pareto front	*aguirre2007*
#plo	proportion of Pareto local optimal (PLO) solutions	*paquete2007*
#slo_avg	average proportion of single-objective local optimal solutions per objective	
podist_avg	average Hamming distance between Pareto optimal solutions	*liefooghe2013*
podist_max	maximal Hamming distance between Pareto optimal solutions (diameter of the Pareto set)	*knowles2003*
po_ent	entropy of binary variables from Pareto optimal solutions	*knowles2003*
fdc	fitness-distance correlation in the Pareto set (Hamming dist. in solution space vs. Manhattan dist. in objective space)	*knowles2003*
#cc	proportion of connected components in the Pareto graph	*paquete2009*
#sing	proportion of isolated Pareto optimal solutions (singletons) in the Pareto graph	*paquete2009*
#lcc	proportional size of the largest connected component in the Pareto graph	*verel2011*
lcc_dist	average Hamming distance between solutions from the largest connected component	
lcc_hv	proportion of hypervolume covered by the largest connected component	
#fronts	proportion of non-dominated fronts	*aguirre2007*
front_ent	entropy of the non-dominated front's size distribution	

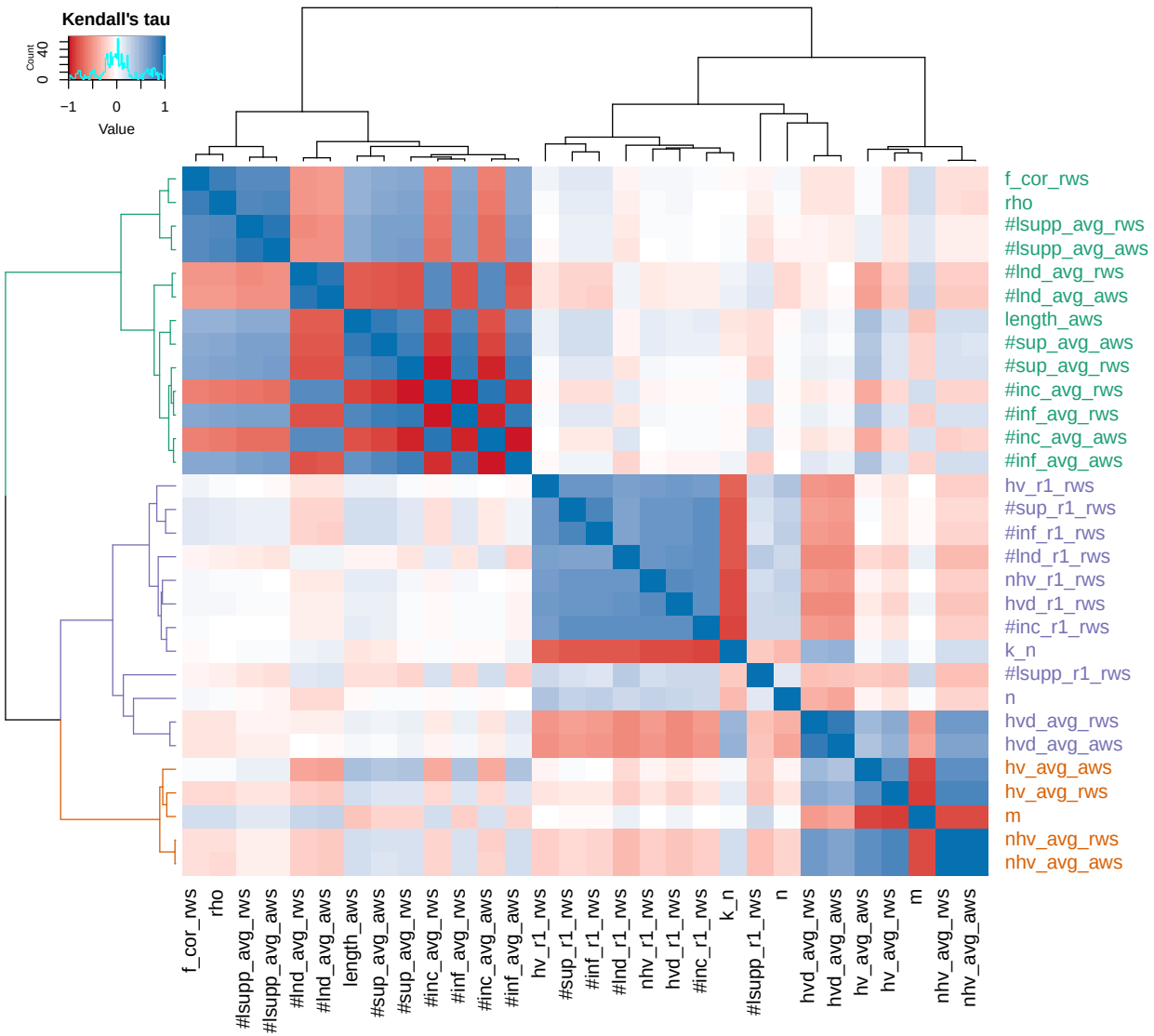
Summary of problem features (2)

	LOCAL FEATURES FROM RANDOM WALK **sampling** (rws) (17)	
hv_avg_rws	average (single) solution's hypervolume-value	
hv_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume-values	*liefooghe2013*
hvd_avg_rws	average (single) solution's hypervolume difference-value	
hvd_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume difference-values	*liefooghe2013*
nhv_avg_rws	average neighborhood's hypervolume-value	
nhv_r1_rws	first autocorrelation coefficient of neighborhood's hypervolume-value	
#lnd_avg_rws	average proportion of locally non-dominated solutions in the neighborhood	
#lnd_r1_rws	first autocorrelation coefficient of the proportion of locally non-dominated solutions in the neighborhood	
#lsupp_avg_rws	average proportion of supported locally non-dominated solutions in the neighborhood	
#lsupp_r1_rws	first autocorrelation coefficient of the proportion of supported locally non-dominated solutions in the neighborhood	
#inf_avg_rws	average proportion of neighbors dominated by the current solution	
#inf_r1_rws	first autocorrelation coefficient of the proportion of neighbors dominated by the current solution	
#sup_avg_rws	average proportion of neighbors dominating the current solution	
#sup_r1_rws	first autocorrelation coefficient of the proportion of neighbors dominating the current solution	
#inc_avg_rws	average proportion of neighbors incomparable to the current solution	
#inc_r1_rws	first autocorrelation coefficient of the proportion of neighbors incomparable to the current solution	
f_cor_rws	estimated correlation between the objective values	
	LOCAL FEATURES FROM ADAPTIVE WALK **sampling** (aws) (9)	
hv_avg_aws	average (single) solution's hypervolume-value	
hvd_avg_aws	average (single) solution's hypervolume difference-value	
nhv_avg_aws	average neighborhood's hypervolume-value	
#lnd_avg_aws	average proportion of locally non-dominated solutions in the neighborhood	
#lsupp_avg_aws	average proportion of supported locally non-dominated solutions in the neighborhood	
#inf_avg_aws	average proportion of neighbors dominated by the current solution	
#sup_avg_aws	average proportion of neighbors dominating the current solution	
#inc_avg_aws	average proportion of neighbors incomparable to the current solution	
length_aws	average length of Pareto-based adaptive walks	*verel2011*

Pairwise feature association (enumerable instances)



Pairwise feature association (large-size instances)



Same association between features from small to large size instances

Experimental setup for large-size instances

Large-size *ρMNK-landscapes*, constrained random LHS DOE

- problem size $n \in [64, 256]$
- problem non-linearity $k \in [0, 8]$
- number of objectives $m \in [2, 5]$
- objective correlation $\rho \in [-1, 1]$, $\rho > \frac{-1}{m-1}$

1 000 problem instances overall

GSEMO and IPLS *algorithms*

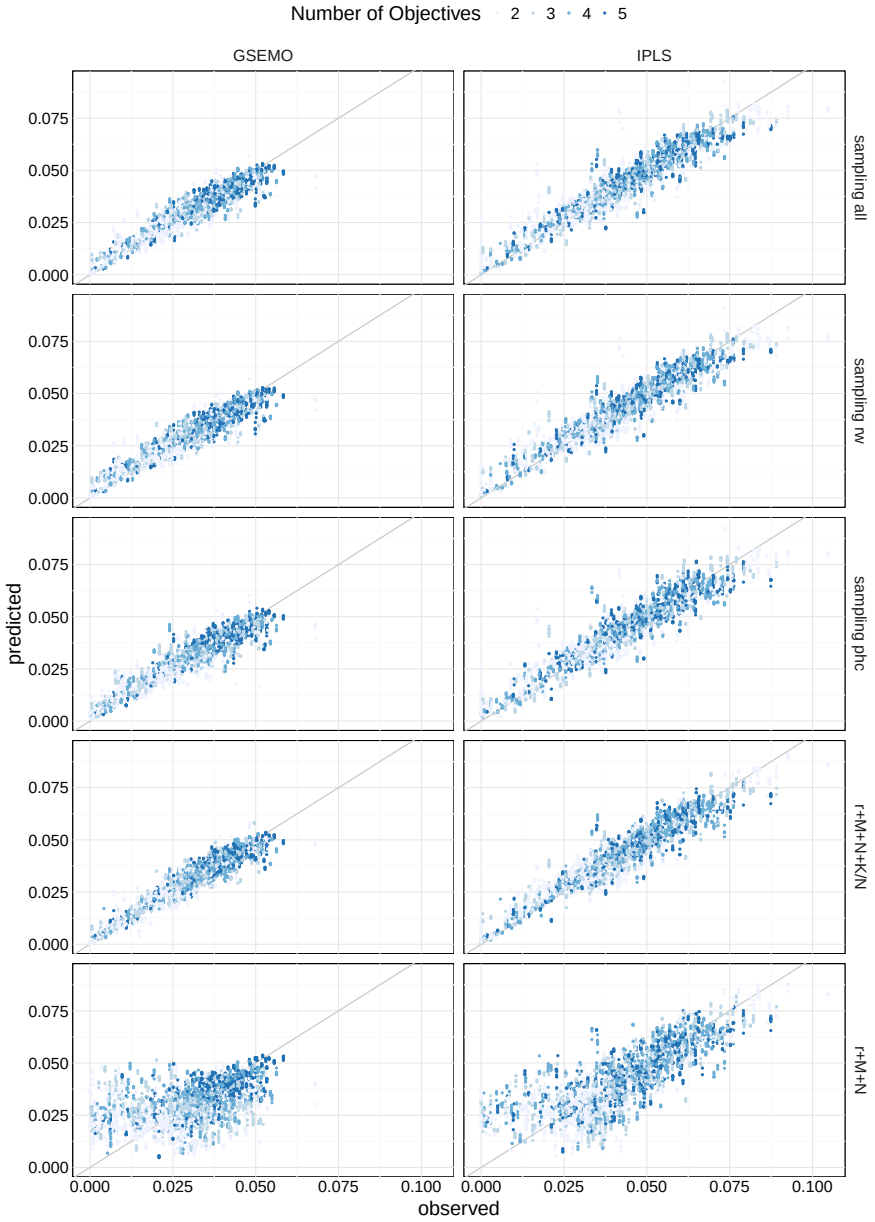
- 30 independent runs per instance
- Fixed budget of 100 000 evaluation calls
- epsilon approximation ratio to best-found non-dominated set

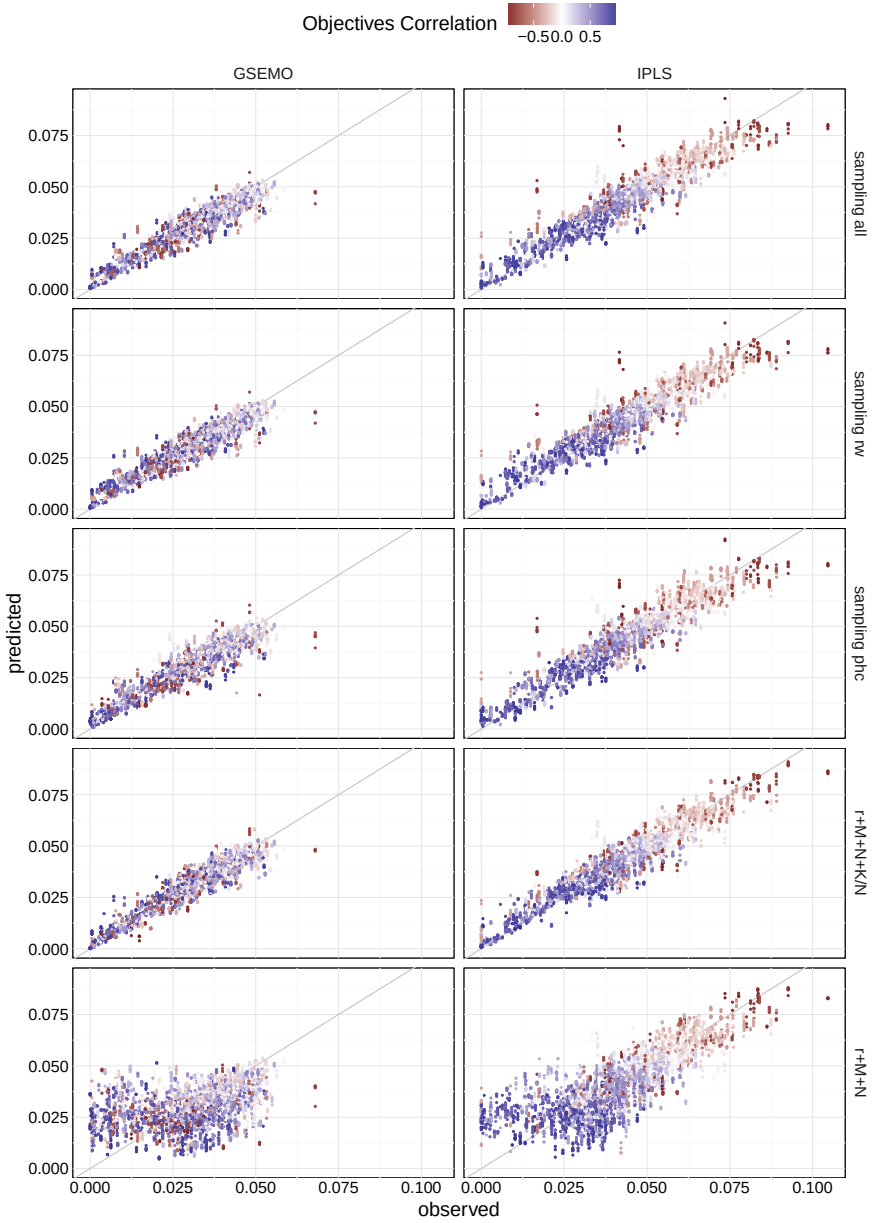
Prediction accuracy

cross validation with repeated subsampling, 50 iterations, 90/10 split

feature set	MAE	MSE	R^2	rank
		GSEMO		
all	**0.003049**	**0.000017**	**0.891227**	1
sampling all	0.003152	**0.000018**	**0.883909**	1.3
sampling rws	0.003220	0.000019	0.878212	2
sampling aws	0.003525	0.000023	0.854199	3
ρ+m+n+k/n	**0.003084**	**0.000017**	**0.892947**	1
ρ+m+n	0.009062	0.000148	0.065258	4
m+n	0.010813	0.000206	-0.303336	5
		IPLS		
all	**0.004290**	**0.000034**	**0.886568**	1
sampling all	**0.004359**	**0.000035**	**0.883323**	1
sampling rws	**0.004449**	0.000036	**0.879936**	1.3
sampling aws	0.004663	0.000039	0.871011	2
ρ+m+n+k/n	**0.004353**	**0.000033**	**0.889872**	1
ρ+m+n	0.008415	0.000119	0.600965	3
m+n	0.016959	0.000472	-0.568495	4

Predicted vs observed values (out-of-folds)





Portfolio accuracy

cross validation with repeated subsampling, 50 iterations, 90/10 split

Portfolio : { GSEMO , IPLS }

feature set	error rate	rank
all	**0.0128**	1
sampling all	**0.0138**	1
sampling rws	**0.0150**	1
sampling aws	**0.0144**	1
ρ+m+n+k/n	**0.0134**	1
ρ+m+n	0.0824	2
m+n	0.1328	3
const=GSEMO	0.0880	
const=IPLS	0.7250	

Conclusions

Fitness landscape to :

- Understand problem structure from the point of view of local search
- Gain knowledge about problem difficulty
- Explain and predict algorithm performance
- Select algorithm

Local optima network :

- variable aggregation principle + complex network analysis

Multi-objective fitness landscape :

- Relevant features

Perspectives / open issues

- Improving the configuration of algorithms using fitness landscape
- Design cheap features
- Large portfolio of multi-objective algorithms
- Theoretical analysis of LON :
  compute the number of local optima, bassin, etc.
- Fitness landscape for population-based algorithm
- Fitness landscape for continuous and multi-objective optimization
- Extend elementary landscape for another differential equation
- Which "**aggregation of variables**" shows relevant properties of the optimization problem according to the local search heuristic ?

$$\begin{tikzcd} X \arrow[r, "op"] \arrow[d, "p"'] & X \arrow[d, "p"] \\ Z \arrow[r, "op_Z"] & Z \end{tikzcd}$$

References I

Khulood Alyahya and Jonathan E Rowe.
Simple random sampling estimation of the number of local optima.
In *International Conference on Parallel Problem Solving from Nature*, pages 932–941. Springer, 2016.

A-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286 :509–512, 1999.

J. P. K. Doye.
The network topology of a potential energy landscape : a static scale-free network.
Phys. Rev. Lett., 88 :238701, 2002.

References II

Fabio Daolio, Marco Tomassini, Sébastien Verel, and Gabriela Ochoa.
Communities of Minima in Local Optima Networks of Combinatorial Spaces.
Physica A : Statistical Mechanics and its Applications, 390(9) :1684 – 1694, July 2011.

Josselin Garnier and Leila Kallel.
Efficiency of local search with multiple local optima.
SIAM Journal on Discrete Mathematics, 15(1) :122–141, 2002.

S. A. Kauffman.
The Origins of Order.
Oxford University Press, New York, 1993.

References III

KatherineM. Malan and AndriesP. Engelbrecht.
Fitness landscape analysis for metaheuristic performance prediction.
In Hendrik Richter and Andries Engelbrecht, editors, *Recent Advances in the Theory and Application of Fitness Landscapes*, volume 6 of *Emergence, Complexity and Computation*, pages 103–132. Springer Berlin Heidelberg, 2014.

Peter F. Stadler.
Landscapes and their correlation functions.
J. Math. Chem., 20 :1–45, 1996.

References IV

P. F. Stadler.
Fitness landscapes.
In M. Lässig and Valleriani, editors, *Biological Evolution and Statistical Physics*, volume 585 of *Lecture Notes Physics*, pages 187–207, Heidelberg, 2002. Springer-Verlag.

Frank H Stillinger and Thomas A Weber.
Packing structures and transitions in liquids and solids.
Science(Washington, DC), 225(4666) :983–9, 1984.

E. D. Weinberger.
Correlated and uncorrelatated fitness landscapes and how to tell the difference.
In *Biological Cybernetics*, pages 63 :325–336, 1990.

References V

S. Wright.
The roles of mutation, inbreeding, crossbreeding, and selection in evolution.
In *Proceedings of the Sixth International Congress of Genetics 1*, pages 356–366, 1932.

D. J. Watts and S. H. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393 :440–442, 1998.