technische universität
dortmund

Two-layered Surrogate Modeling for Tuning Optimization Metaheuristics

Günter Rudolph, **Mike Preuss & Jan Quadflieg**

Lehrstuhl für Algorithm Engineering

Fakultät für Informatik

TU Dortmund

enbis emse2009
G2I Centre / 3MI Department
SAINT-ETIENNE / FRANCE
1 / 2 / 3 JULY
Design and Analysis
of Computer Experiments

Outline

- Introduction: Main Goal and Ideas
- Layer 1: Model-assisted Evolution Strategy (MAES)
- Layer 2: Sequential parameter optimization (SPO)
- Proof of Principle: Benchmark Problems
- The Real Thing: Optimization of Ship Propulsion System (Linearjet)
- Conclusions

Main Goal

development of an efficient method for

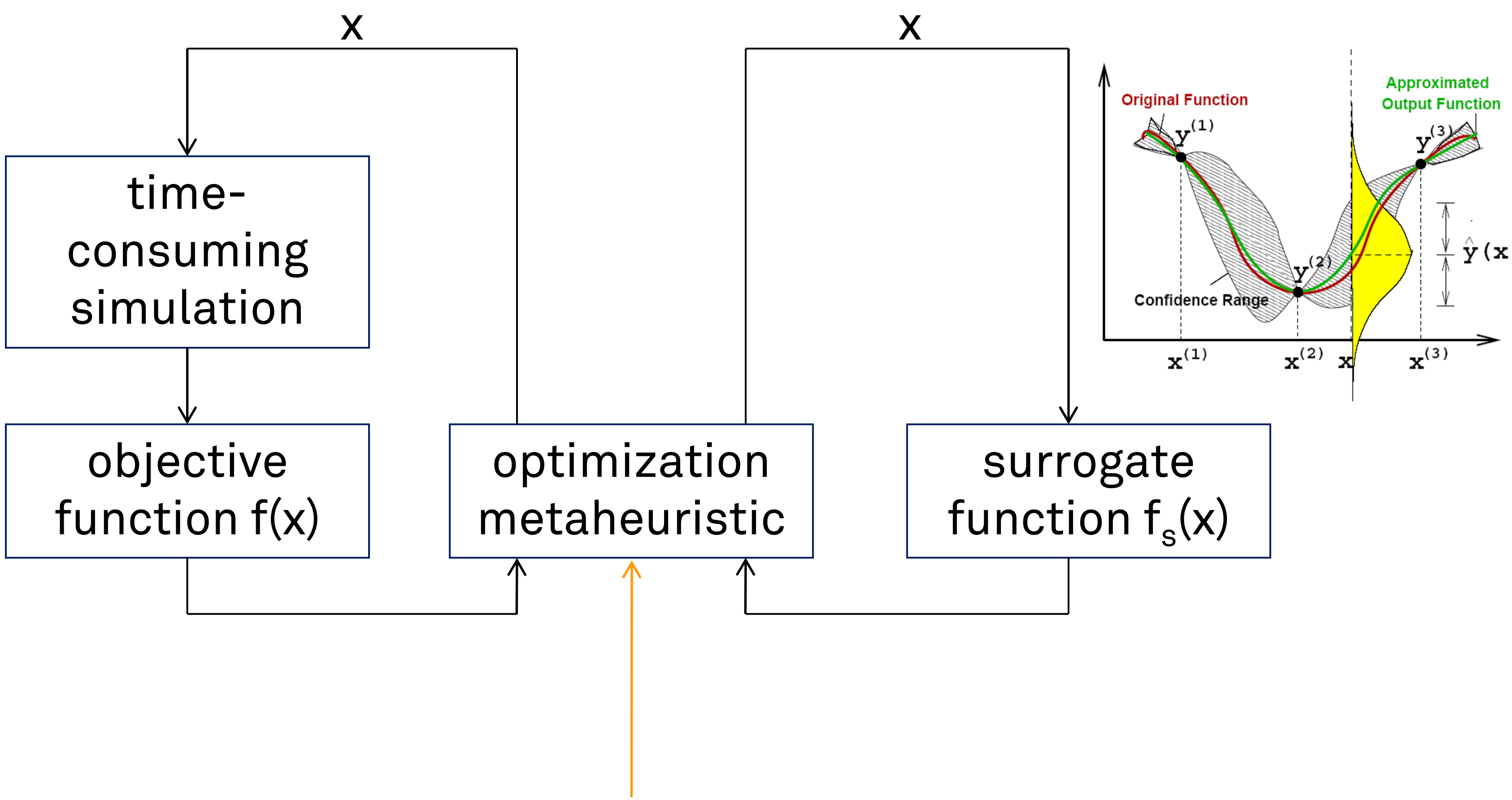
finding good parameterization of a stochastic optimization algorithm

applied to problems with time-consuming objective function

$\Rightarrow$ we do not focus on optimizing objective function, here

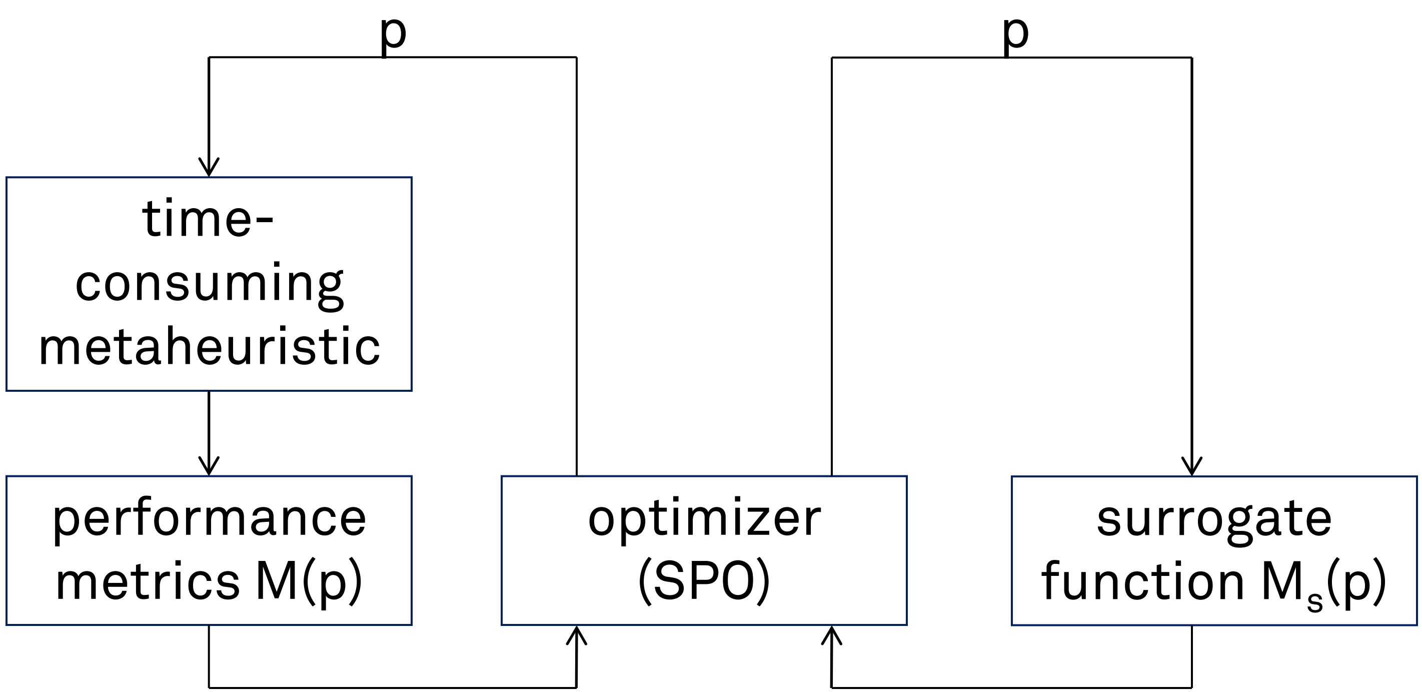
$\Rightarrow$ rather, identify good parameterization of metaheuristic

before spending time, effort, money etc. on optimization of true problem



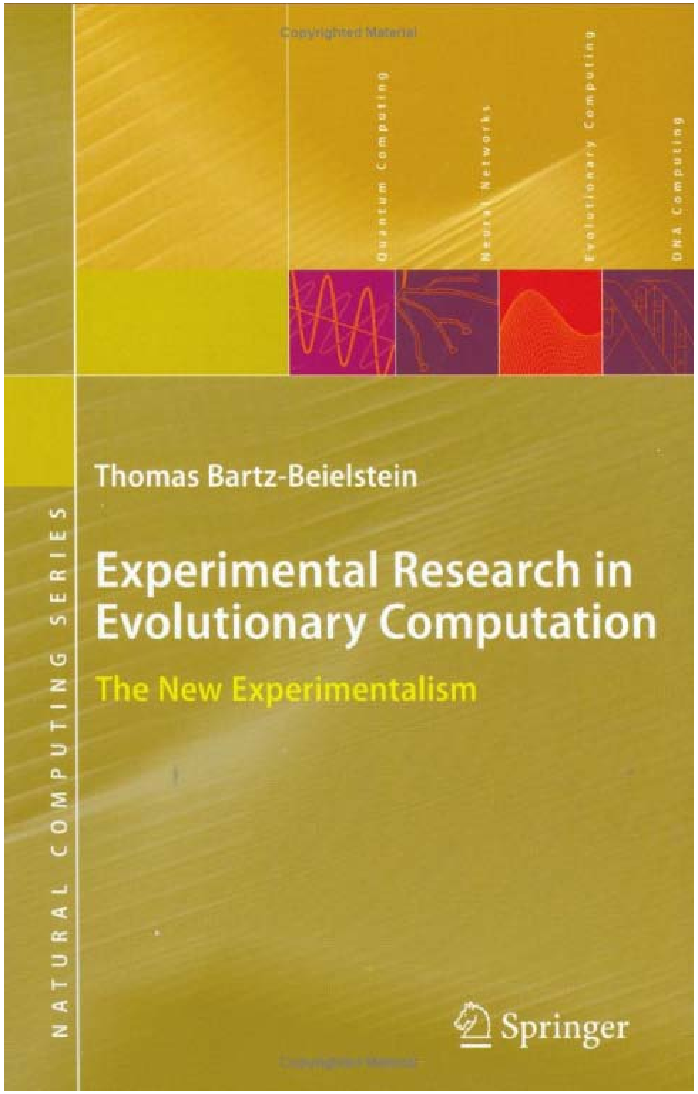
good parameterization of metaheuristic?

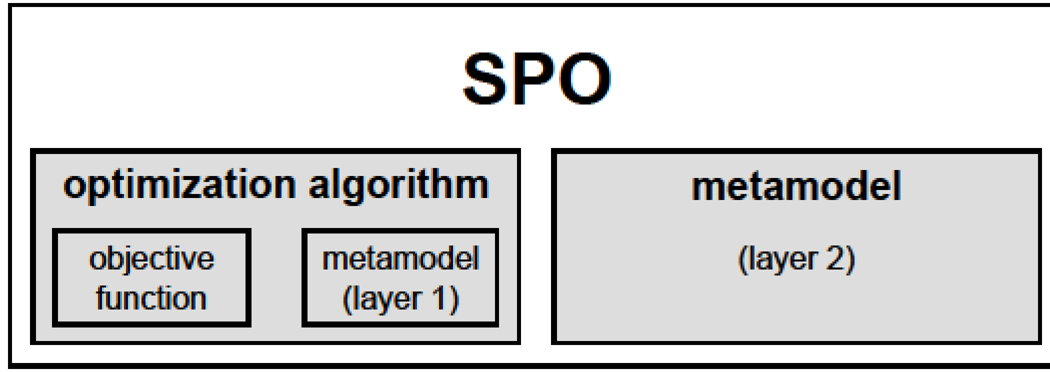
Scenario – Part II



$\Rightarrow$ optimize parameters p of metaheuristic
$\Rightarrow$ result M(p) is a random variable!
$\Rightarrow$ kind of noisy optimization
$\Rightarrow$ repeated evaluation & averaging (roughly)

Two Layers of Meta- / Surrogate Models





Assumptions:

1. Parameter tuning easier than solving optimization problem
2. Rough approximation in layer 1 good enough to allow for tuning metaheuristic

ad 1) fewer parameters ($\approx$ 5) and prior knowledge about metaheuristic

ad 2) to be tested experimentally

Phase 1

Minimal space filling design in parameter space
Run metaheuristic for each design p
Yields pair { p, M(p) } per run and many pairs { x, f(x) } over all runs
Pairs { x, f(x) } used to build 1st layer surrogate model $f_s(x)$

Phase 2

SPO uses pairs { p, M(p) } to build 2nd layer surrogate model $M_s(p)$

```
repeat
```

  SPO optimizes parameters p on $M_s(p)$
  Yields first candidate p*
  Validation runs on f(.) with parameterization p*
  Yields pairs { x, f(x) } $\rightarrow$ update surrogate model $f_s(x)$
  Yields mean pair { p*, M(p*) } $\rightarrow$ update surrogate model $M_s(p)$

`until` resources exhausted

Algorithm 1 $(\mu + \nu < \lambda)$-MAES

$t \leftarrow 0$
$P_t \leftarrow \text{init}()$ /* $P_t \in \mathbb{S}^\mu$: Set of solutions */
evaluate P_t and insert results to database D
while $t < t_{\max}$ **do**
  $G_t \leftarrow \text{generate}(P_t)$ /* Generate λ variations */
  evaluate G_t with meta-model derived from D
  choose set of (maximal) ν promising solutions $Q_t \subseteq G_t$
  evaluate Q_t and update database D with results
  $P_{t+1} \leftarrow \text{select}(Q_t \cup P_t)$ /* Rank and select μ best */
  $t \leftarrow t + 1$
end while

Parameters: μ , λ , k, σ , τ ($\nu = \lambda / 2$)

also testing benefit of external databases
→ initial sizes: 0, 1000, 2000 pairs

surrogate model:
ordinary kriging

- Latin hypercube design in ***parameter space*** (here: 25 with 4 repeats)

- Global ordinary kriging model to predict promising regions

- Deploys expected improvement criterion of EGO

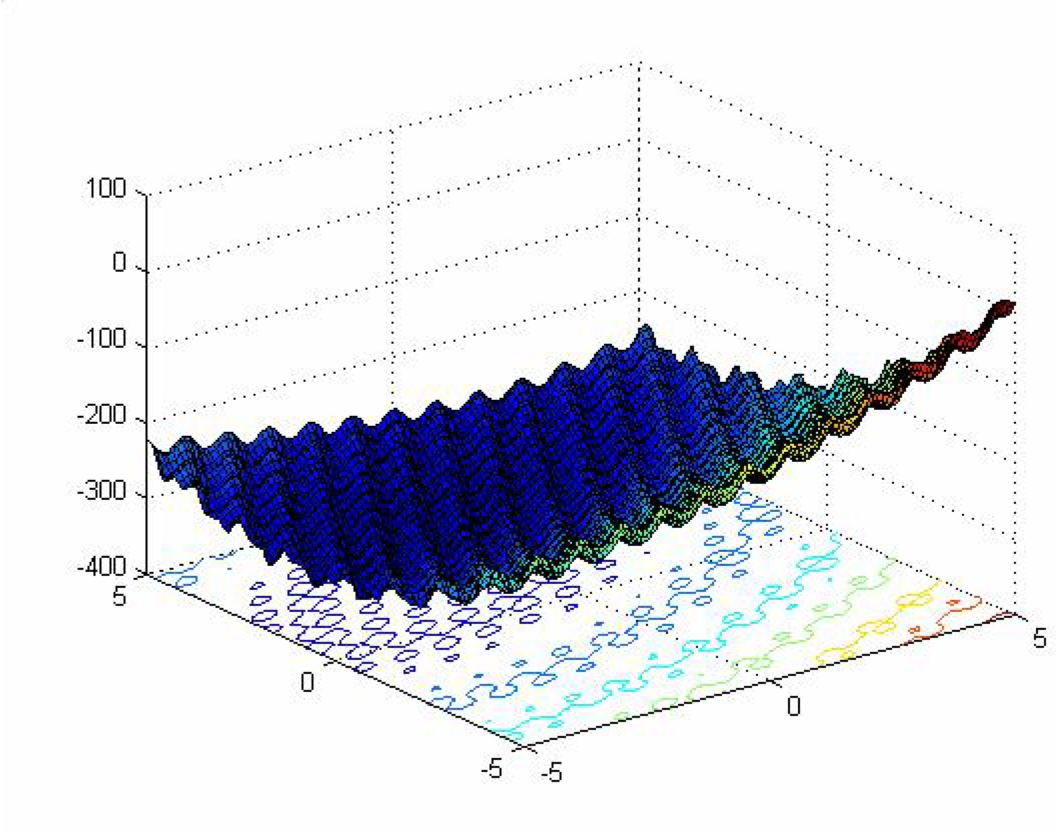
- Considers predicted error and function value

Non-deterministic answers: increasing number of repeats

Total budget of algorithm runs: 500 (here)

taken from
IEEE CEC'05 benchmark

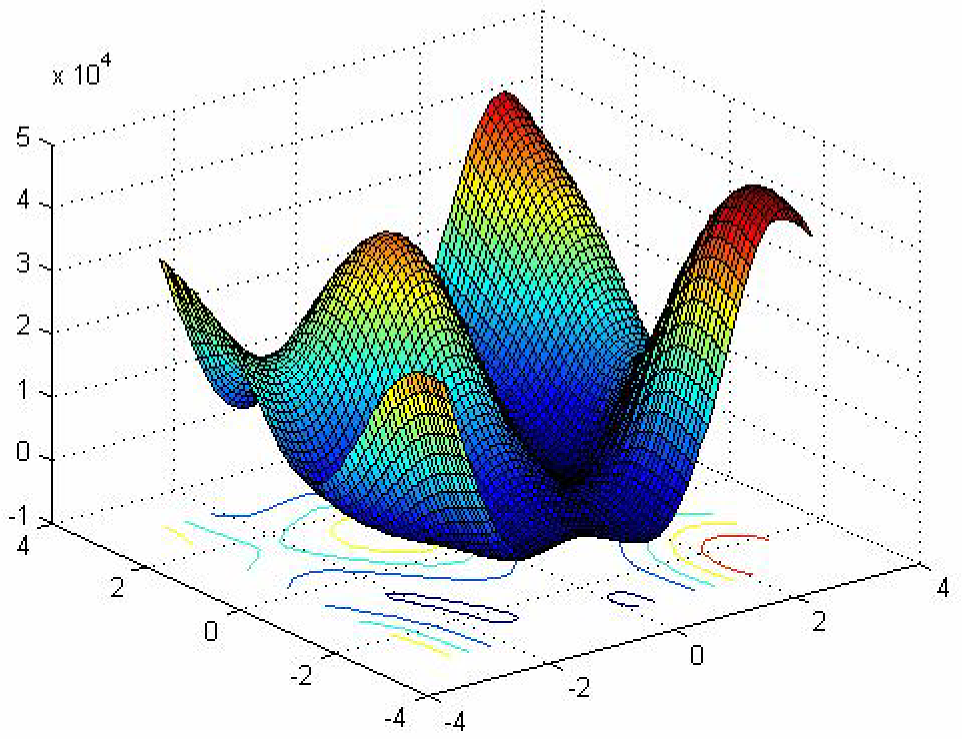
$f_{10}(x)$ Rotated Rastrigin dimensions: 2 and 10



taken from
IEEE CEC‘05 benchmark

$f_{12}(x)$ Schwefel 2.13

dimensions: 2 and 10



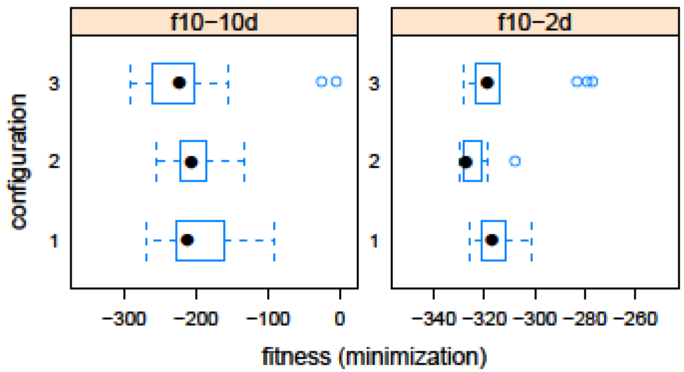
Tuning on 1st layer successful?

20 runs for each database size $\in \{ 0, 1000, 2000 \}$

initial: $\sigma = 0.15$, $\tau = 1.0$, $k = 10$, $\mu = 1$, $\lambda = 5$

Rotated Rastrigin

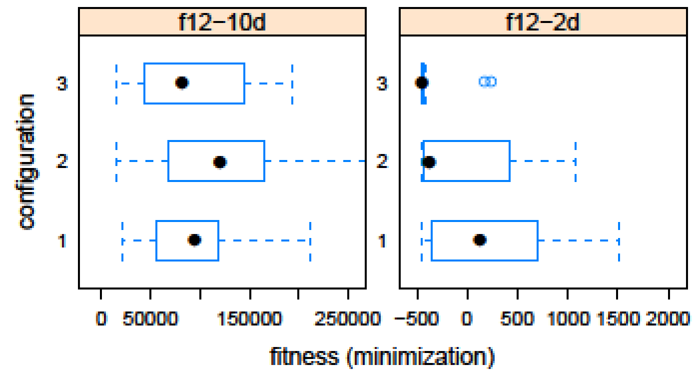
Schwefel 2.13

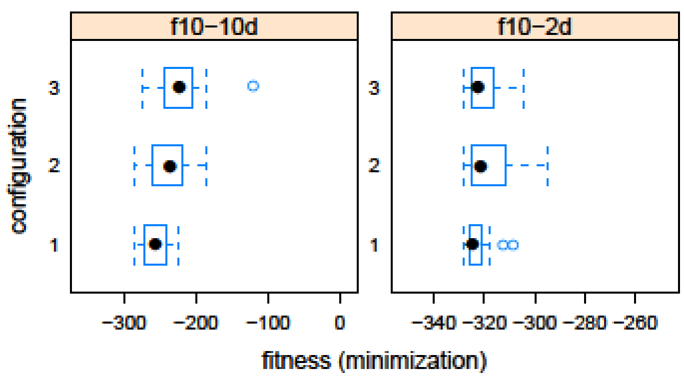


D = 2000

D = 1000

D = 0

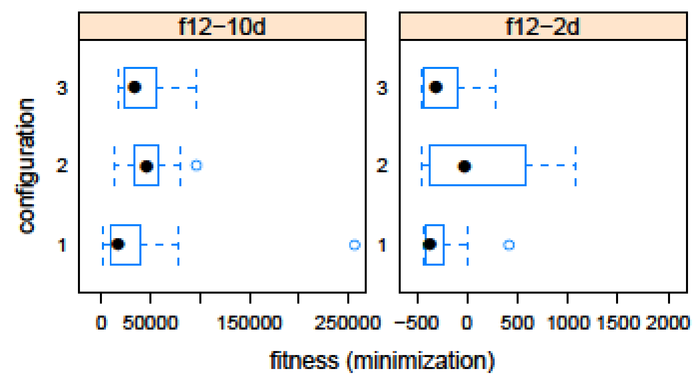




D = 2000

D = 1000

D = 0



Tuning on 1st layer successful?

Standard initial and tuned parameters of MAES

	σ_{init}	τ	κ	μ	λ
standard	0.15	1.0	10	1	5
f10 2d SPO(O)	0.088	0.753	6	3	13
f10 2d SPO(1000)	0.338	1.662	15	3	14
f10 10d SPO(O)	0.400	0.525	19	5	11
f10 10d SPO(1000)	0.415	1.525	20	5	12
f12 2d SPO(O)	0.055	0.114	20	5	14
f12 2d SPO(2000)	0.3974	0.696	29	4	14
f12 10d SPO(O)	0.017	0.173	23	5	13
f12 10d SPO(1000)	0.208	0.210	10	4	11

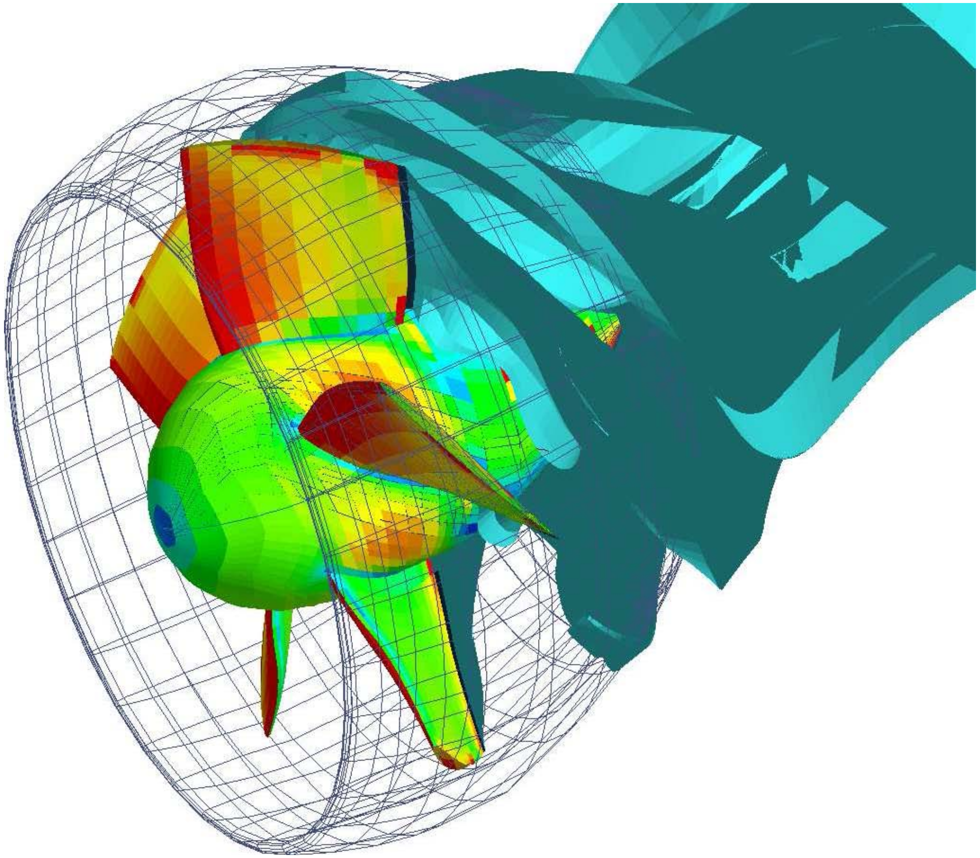
Tuning on 1st layer successful?

p-values of Wilcoxon rank-sum test at level 0.05

between 20 validation runs of different parameter sets

	f10 2d	f10 10d	f12 2d	f12 10d
standard : SPO(O)	10^{-3}	10^{-5}	0.010	10^{-5}
standard : SPO(1000)	0.134	0.005	0.799	10^{-4}
standard : SPO(2000)	0.059	0.277	0.013	10^{-4}
SPO(O) : SPO(1000)	0.142	0.024	0.021	0.006
SPO(O) : SPO(2000)	0.883	10^{-4}	0.038	0.013
SPO(1000):SPO(2000)	0.142	0.102	0.883	0.369

Linear-jet Engine



15 design variables:
- lengths
- thicknesses
- angles

basic fluid dynamic simulation
needs 3 minutes

full CFD simulation
needs 8 hours in parallel

objective:
minimum cavitation
at a predefined efficiency

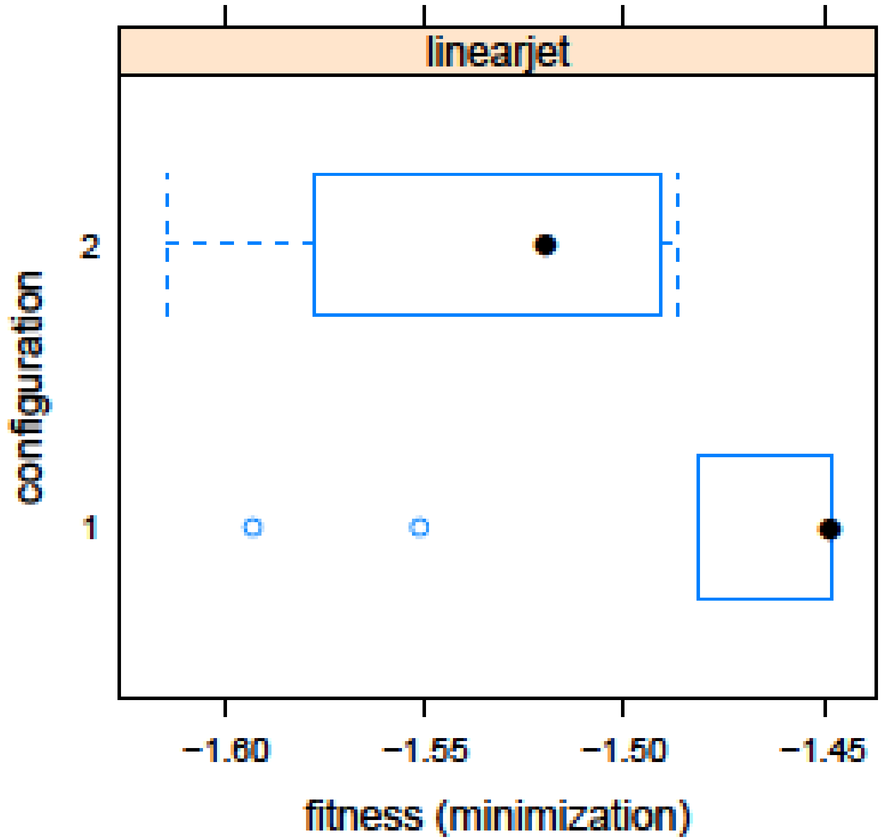
caviatation =
emergence of vacuum bubbles caused by extreme pressure differences
due to high accelerations of the water (causes damage and noise)

database of 2000 points from previous runs
used to create ordinary kriging model

SPO: run lengths of 300 evaluations on true problem

MAES runn 9 times with best parameterization found

note: a single run needs 12 to 24 hours on modern PC



Good values: ≈ -1.6

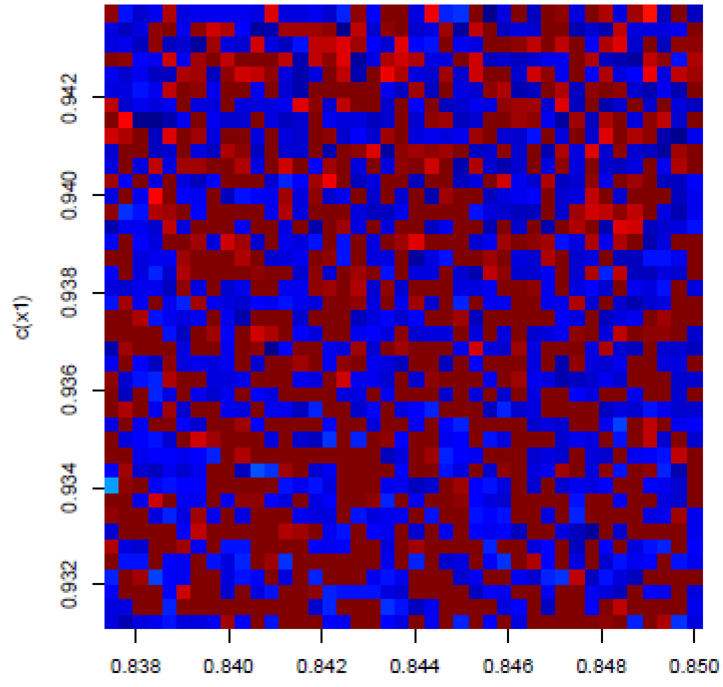
different with p-value 0.02

Parameterization found

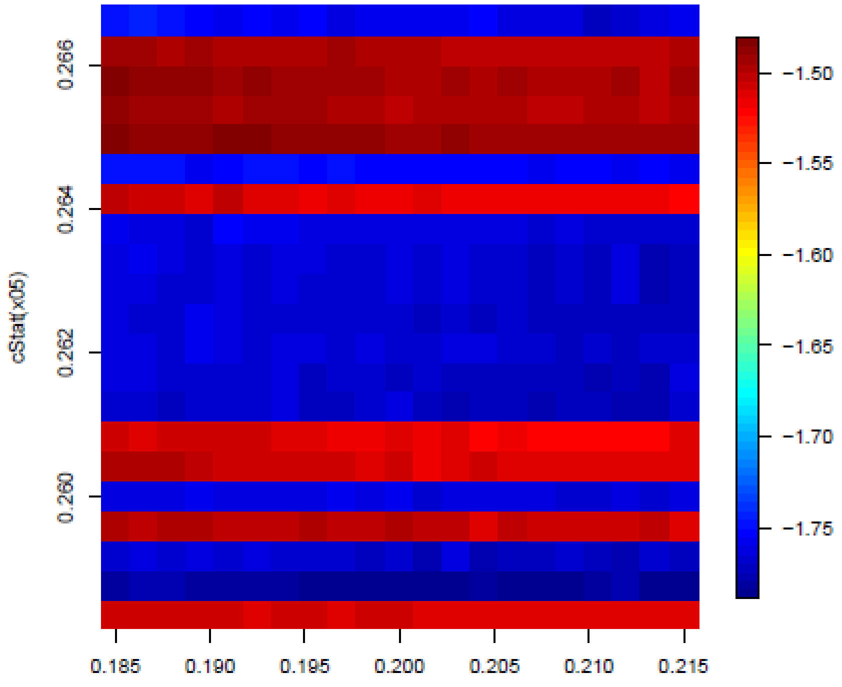
	σ_{init}	τ	κ	μ	λ
standard	0.15	1.0	10	1	5
f10 2d SPO(O)	0.088	0.753	6	3	13
f10 2d SPO(1000)	0.338	1.662	15	3	14
f10 10d SPO(O)	0.400	0.525	19	5	11
f10 10d SPO(1000)	0.415	1.525	20	5	12
f12 2d SPO(O)	0.055	0.114	20	5	14
f12 2d SPO(2000)	0.3974	0.696	29	4	14
f12 10d SPO(O)	0.017	0.173	23	5	13
f12 10d SPO(1000)	0.208	0.210	10	4	11
→ linearjet SPO(2000)	0.170	0.794	20	4	11

Modelling of objective function needs revision
→ evidently, penalizations lead to rugged response surface

20x20 grid
2 rotor parameters



20x20 grid
uncorrelated parameters



Two-layered surrogate model approach works quite well

... but needs more work

MAES not best choice $\rightarrow$ replace by other metaheuristic

Hypothesis:
works since only main characteristics of true problem must be reflected by surrogate model

$\rightarrow$ theoretical foundation possible?

Future work: integrated / automatic procedure

Questions?

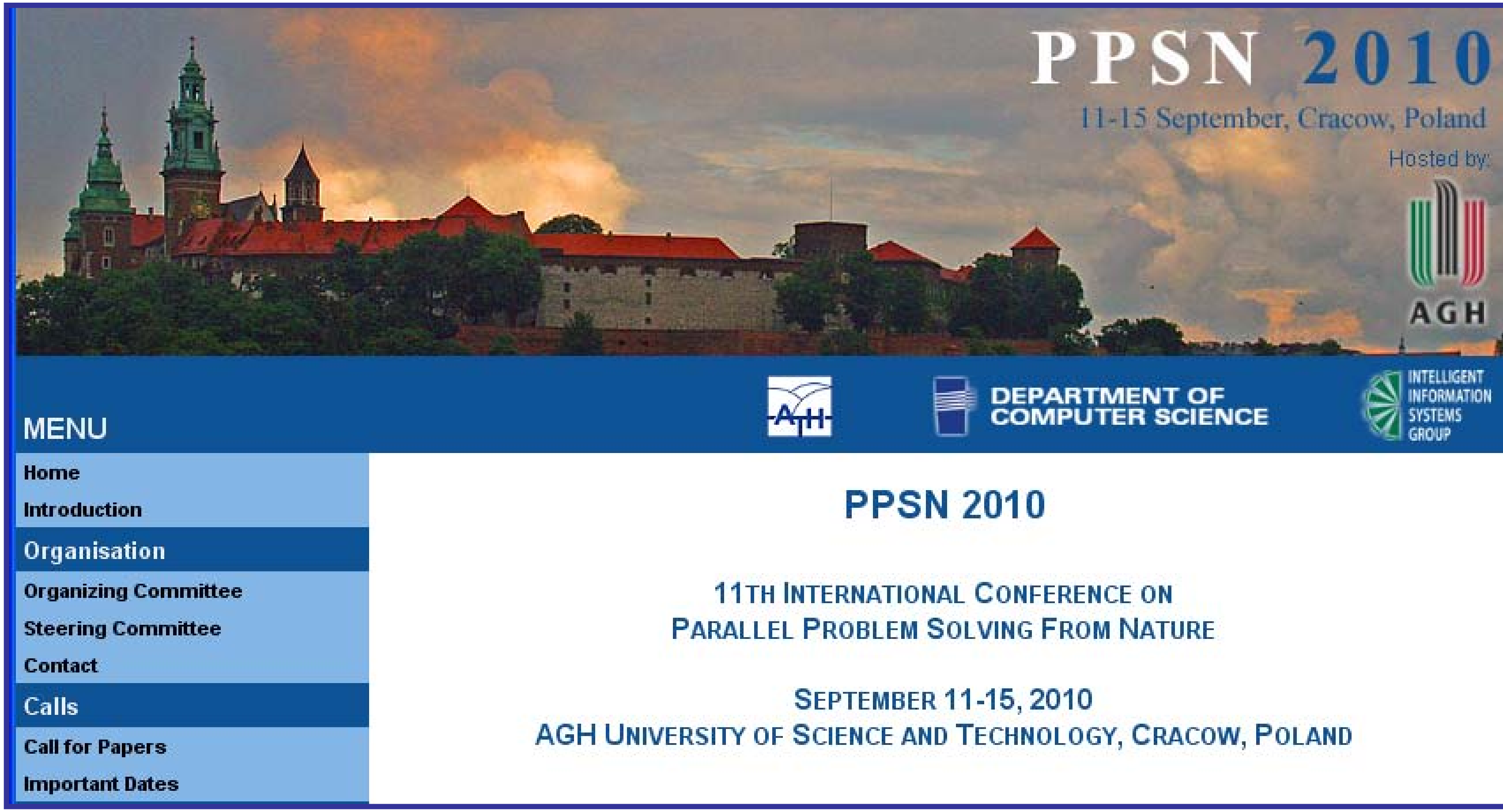
WCCI 2010

2010 IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE

JULY 18-23 CENTRE DE CONVENCIONS INTERNACIONAL DE BARCELONA, BARCELONA, SPAIN

Important Deadlines

- **November 15, 2009** *Competition proposals*
- **December 13, 2009** *Special sessions proposals*
- **December 22, 2009** *Notification of special session acceptance*
- **January 31, 2010** *Paper submission*
- **February 14, 2010** *Tutorial and workshop proposal*



Paper submission: April 5, 2010