Multi-objective hyperparameter optimization with performance uncertainty

Alejandro Morales-Hernández [1,2,*], Inneke Van Nieuwenhuyse [1,2], Gonzalo Nápoles [3]
[1] Core Lab VCCM, Flanders Make, Belgium
[2] Data Science Institute, Hasselt University, Belgium
[3] Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands

International Conference in Optimization and Learning, July 18th-20th, Syracuse, Italy

* Contact to: alejandro.moraleshernandez@uhasselt.be

Multi-objective hyperparameter optimization with performance uncertainty

- Multi-objective hyperparameter optimization (MO-HPO)
- Metamodel-based optimization
- Proposed algorithm: GPR + TPE
- Simulation results and final remarks

Multi-objective hyperparameter optimization with performance uncertainty

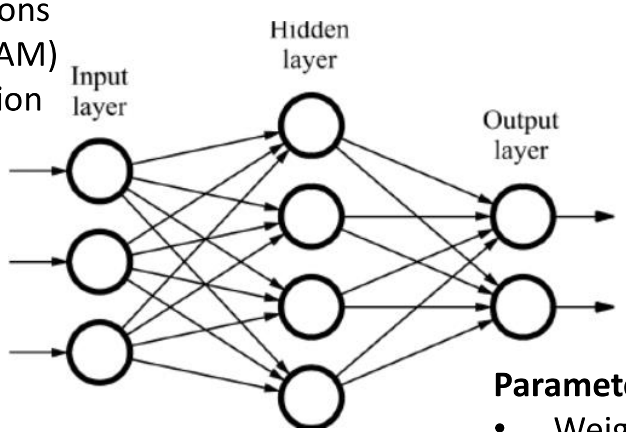
- Multi-objective hyperparameter optimization (MO-HPO)
- Metamodel-based optimization
- Proposed algorithm: GPR + TPE
- Simulation results and final remarks

Hyperparameter:

- Influences the learning process
- Its optimization is not part of the ML algorithm
- Complex domain (numeric, discrete, etc)
- Should be specified before the training phase

Hyperparameter:

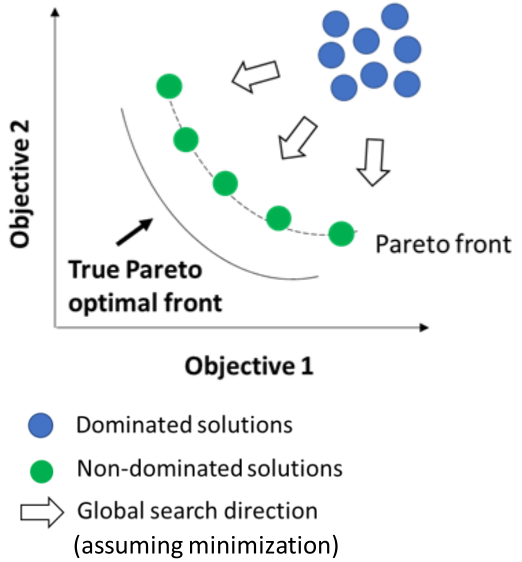
- Number of layers
- Number of neurons
- Solver (SGD, ADAM)
- Activation function
- Learning rate



Parameter:

- Weights

Multi-objective Hyperparameter optimization:



Validation protocol

m performance functions

Validation dataset

$$\lambda^* = \underset{\lambda \in \Lambda}{\operatorname{argmin}}\, V((f_1, \ldots, f_m) | A_\lambda, D_{train}, D_{validation})$$

Optimal set of hyperparameter configurations

Training dataset

The algorithm A with its hyperparameters instantiated to λ is denoted by A_λ

Multi-objective hyperparameter optimization with performance uncertainty

- Multi-objective hyperparameter optimization (MO-HPO)
- Metamodel-based optimization
- Proposed algorithm: GPR + TPE
- Simulation results and final remarks

Replace the expensive black-box functions by others of easy execution
- Gaussian Process Regression (GPR)
- Tree Parzen Estimators (TPE)

Metamodel-based optimization

Gaussian Process Regression (GPR)

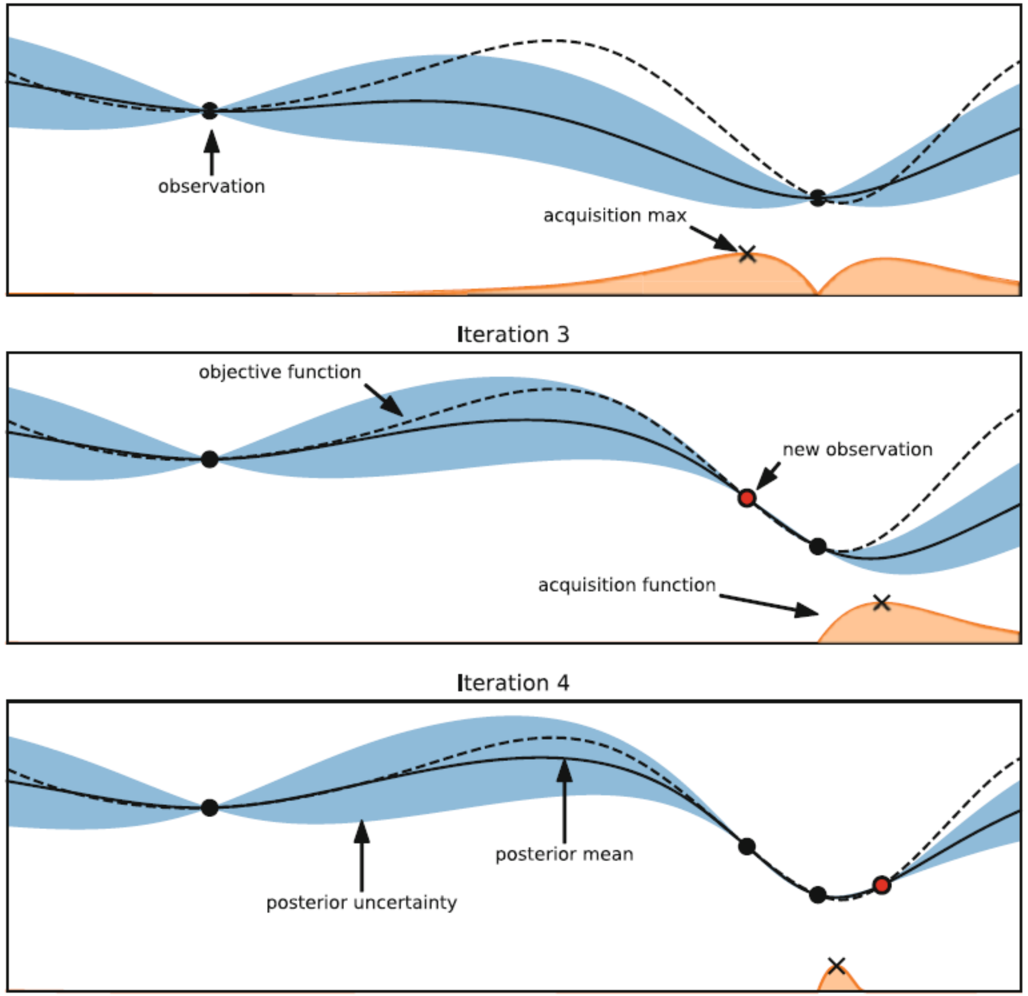


Illustration of Bayesian optimization. The goal is to minimize the dashed line using a Gaussian process surrogate (no noise)

GPR with heteroscedastic noise

Mean of the process

Noise observed in replication *r*

$$y(x) = m(x) + M(x) + \epsilon_r(x)$$

unknown response function

Realization of a Gaussian random field with mean zero

$M(x)$ can be seen as a function that exhibits spatial correlation according to a covariance function

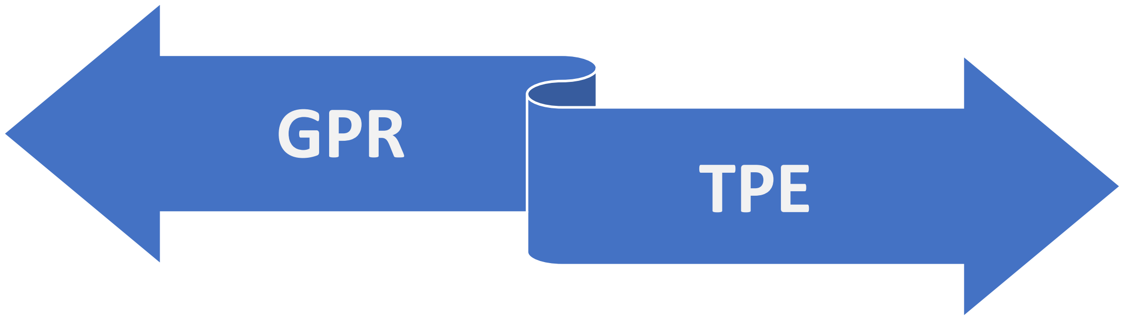
$$Cov\left(y_i, y_j\right) = k(x_i, x_j)$$

- Gaussian kernel
- Mátern kernel

Metamodel-based optimization

Tree Parzen Estimators (TPE), single objective

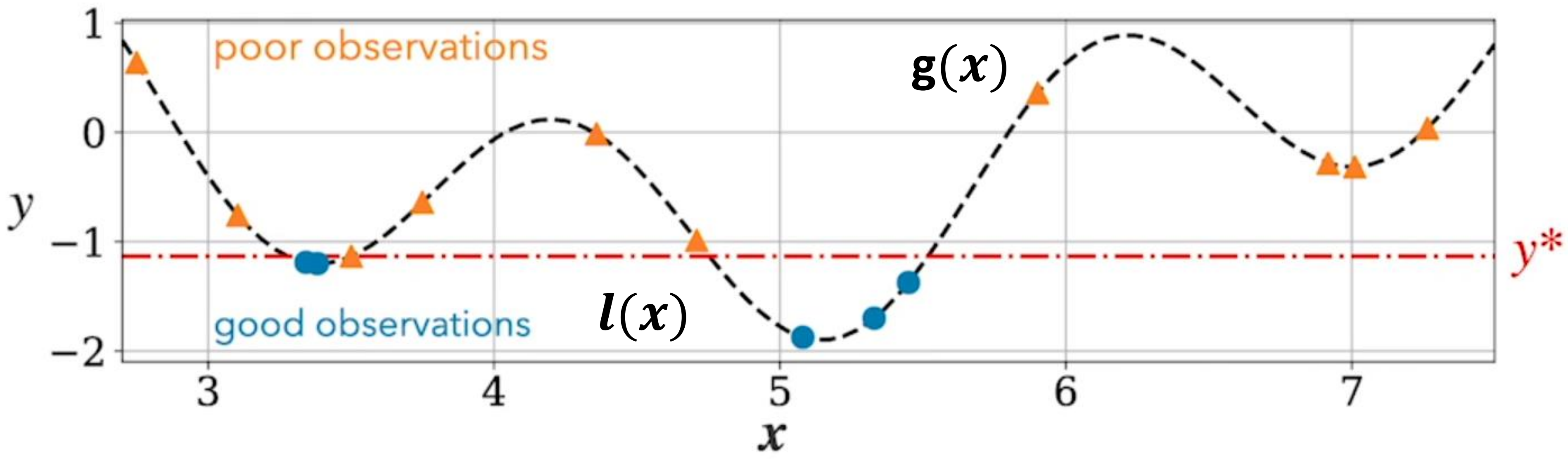
$P(f(x_*)|X,Y)$



$P(x_*|X,Y)$

$$P(x_*|X,Y) = \begin{cases} l(x) & f(x) < y^*, x \in \boldsymbol{X} \\ g(x) & o.w \end{cases}$$

- y^* is some quantile γ of the observed $\boldsymbol{Y}$ values, so that $p(y < y^*) = \gamma$

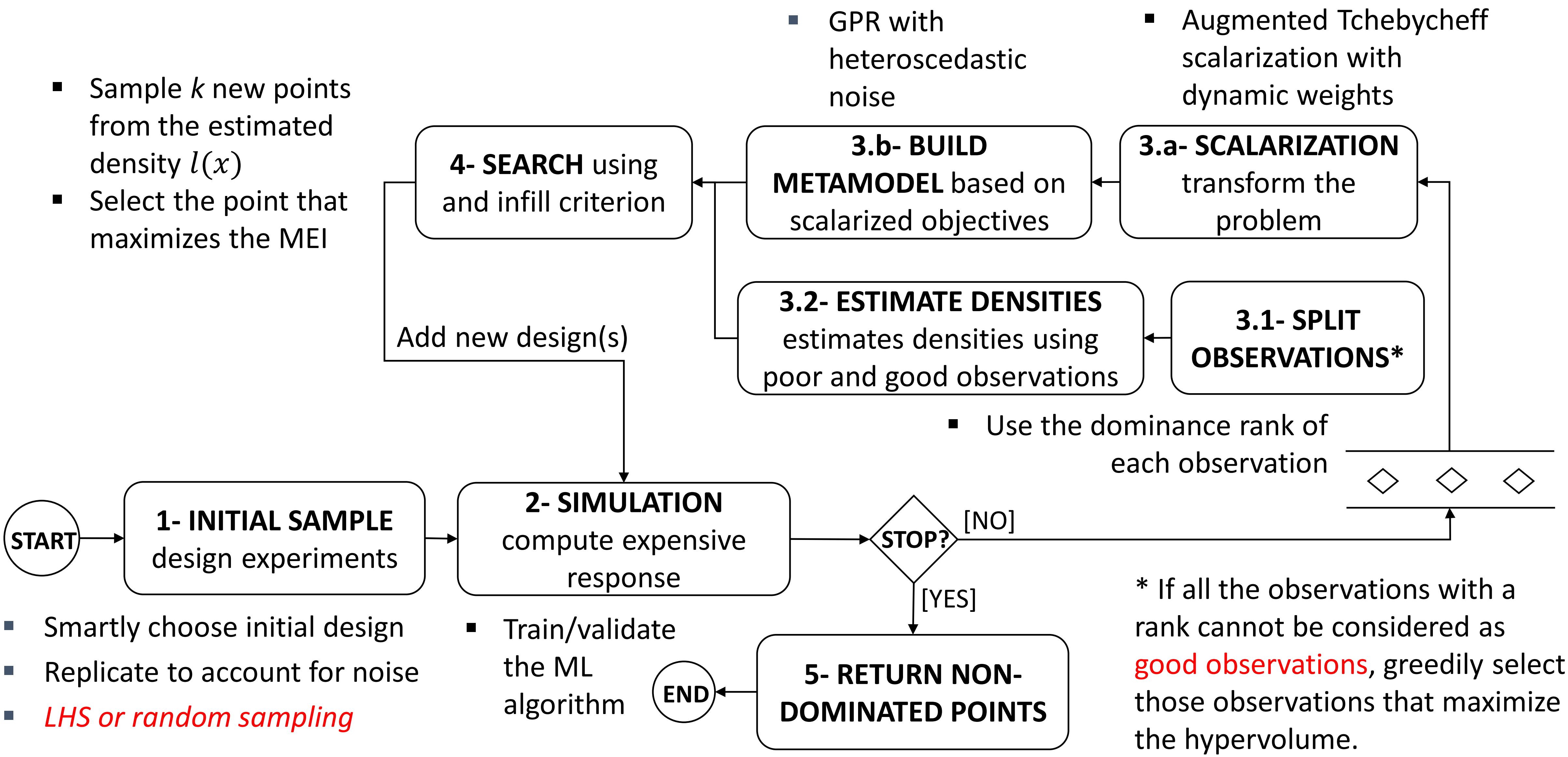


Multi-objective hyperparameter optimization with performance uncertainty

- Multi-objective hyperparameter optimization (MO-HPO)
- Metamodel-based optimization
- Proposed algorithm: GPR + TPE
- Simulation results and final remarks

GPR: handle uncertainty
TPE: sampling strategy

Combining GPR and TPE for multi-objective hyperparameter optimization



Multi-objective hyperparameter optimization with performance uncertainty

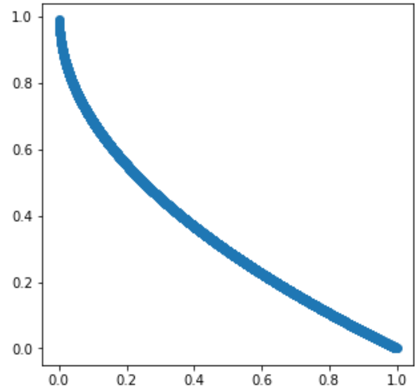
- Multi-objective hyperparameter optimization (MO-HPO)
- Metamodel-based optimization
- Proposed algorithm: GPR + TPE
- Simulation results and final remarks
    - Analytical test functions
    - HPO

Numerical simulations

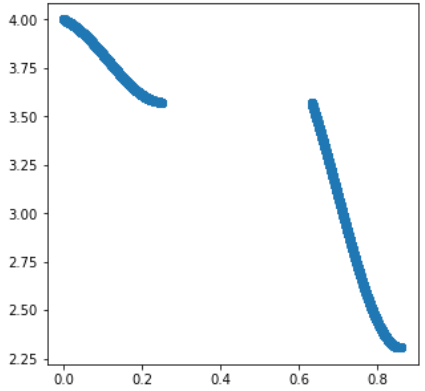
12 OpenML datasets

Experiment 1: Analytical test functions (d=5)

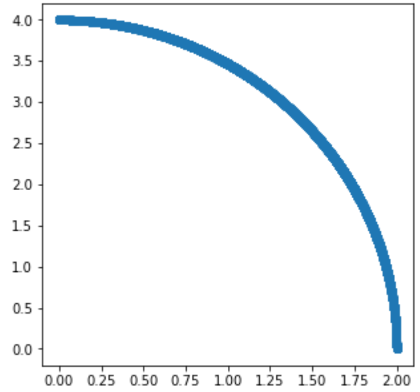
ZDT1



DTLZ7



WFG4



Experiment 2: Hyperparameter optimization (binary classification problem)

Multilayer Perceptron (d=5)		
max_iter	Integer	$[1, 1000]$
neurons	Integer	$[5, 1000]$
lr_init*	Integer	$[10^{-1}, 10^{-6}]$
b1	Real	$[10^{-7}, 1]$
b2	Real	$[10^{-7}, 1]$
activation	Category	$relu$
solver	Category	$adam$
layers	Integer	1

Decision Tree (d=5)		
Max_depth	Integer	$[0, 20]$
mss	Real	$[0, 0.99]$
msl	Integer	$[1, 10]$
max_f	Category	$auto, sqrt, log2$
criterion	Category	$gini, entropy$

Support Vector Machine (d=2)		
C	Real	$[0.1, 2]$
kernel	Category	$linear, poly,$ $rbf, sigmoid$

* Exponent optimization

Performance measures:

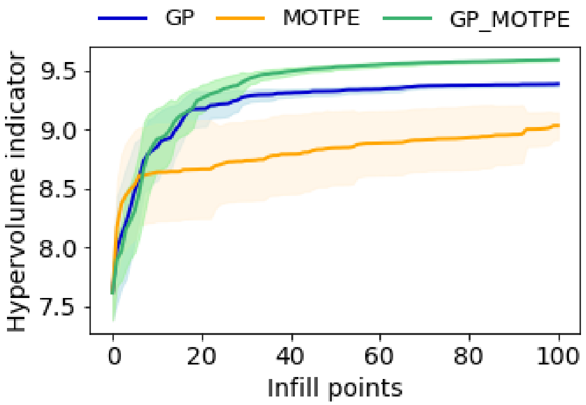
- Minimize error
- Maximize recall

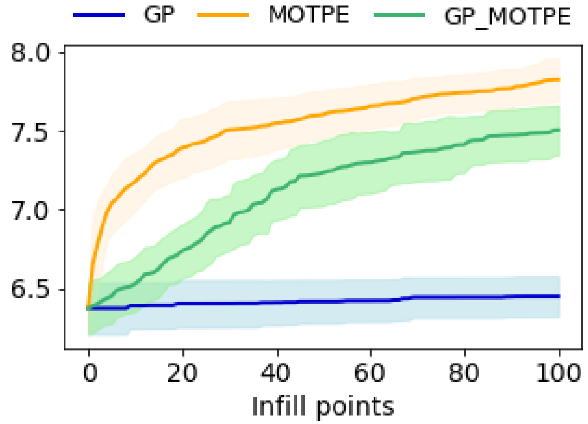
Experimental settings

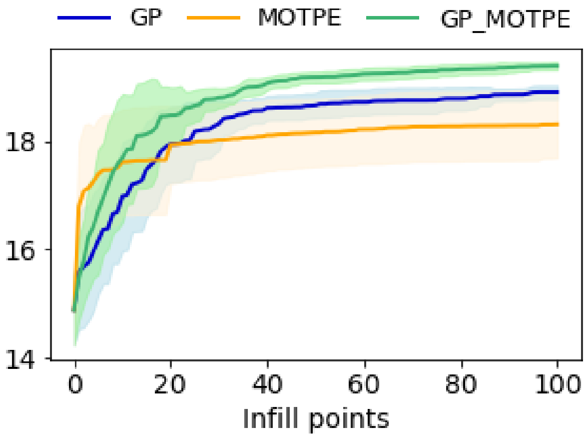
		Algorithm		
Setting	**Problem**	**GPR**	**MOTPE**	**GP_MOTPE**
Design space size	Analytical functions	Latin Hypercube sampling: $11d-1$		
	HPO	Random sampling: $11d-1$		
Replications	Analytical functions	50		
	HPO	10 (*k* value in a cross-validation protocol)		
Iterations	Analytical functions	100		
	HPO			
Acquisition function		MEI	EI_{TPE}	MEI
Acquisition function optimization		PSO	Maximize the acquisition function on a candidate set	
Number of candidates to sample		-	$n_c = 1000, \gamma = 0.3$	
kernel		Gaussian	-	Gaussian

Results

Experiment 1: Analytical test functions (d=5). Hypervolume and Pareto front analysis



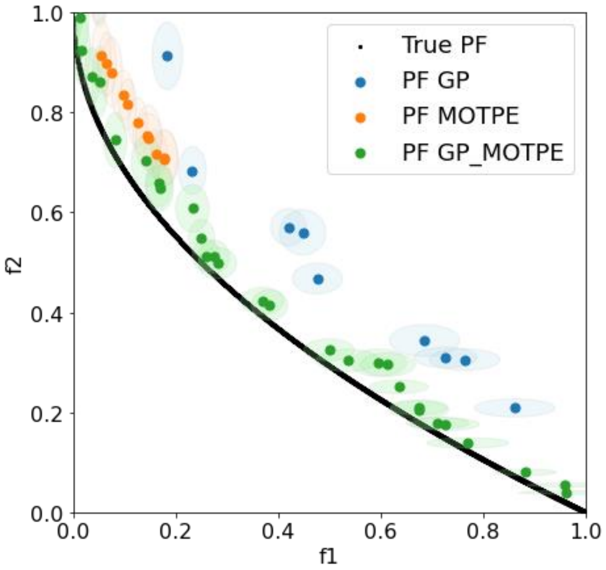


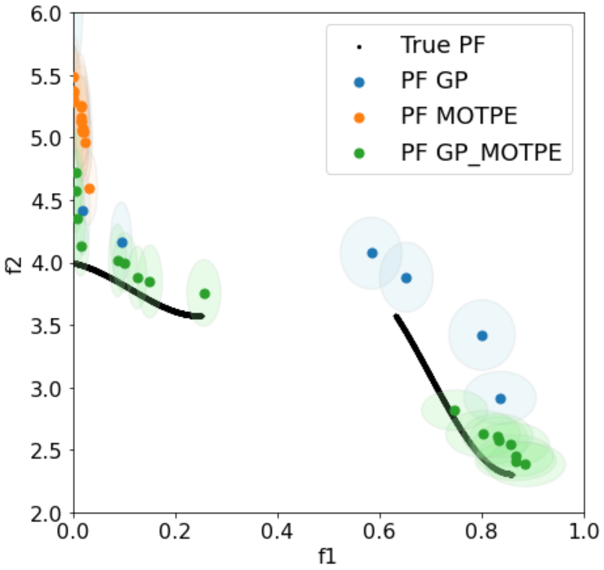


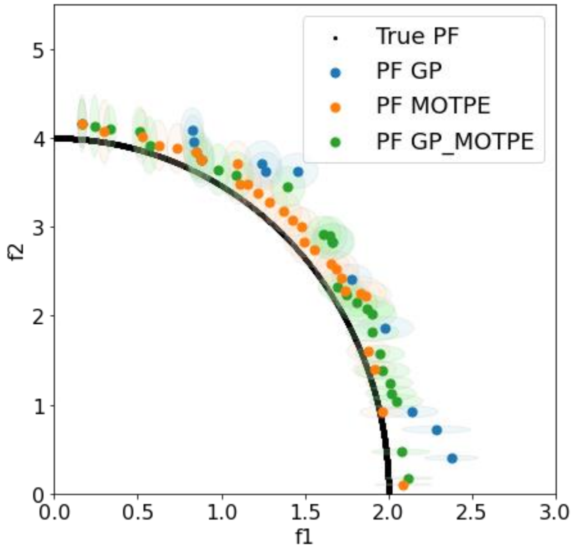
ZDT1
$ref = [1,10]$

DTLZ7
$ref = [1, 23]$

WFG4
$ref = [3, 5]$

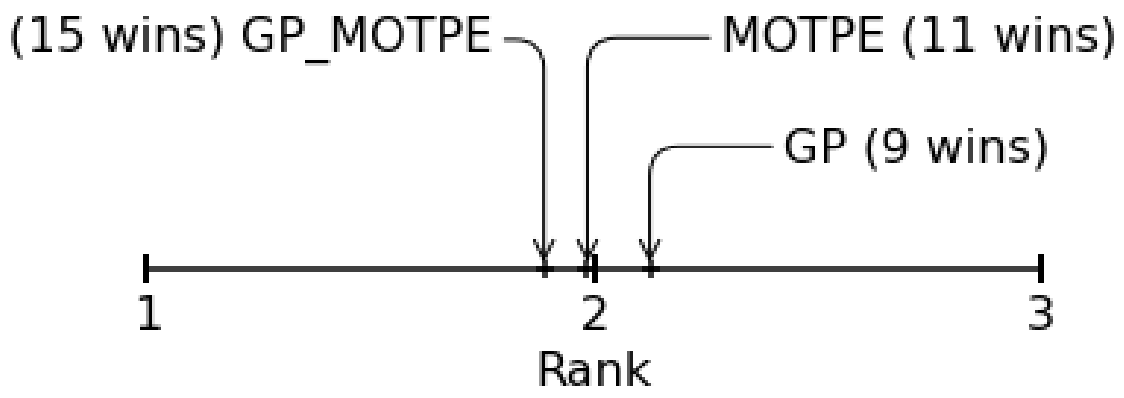


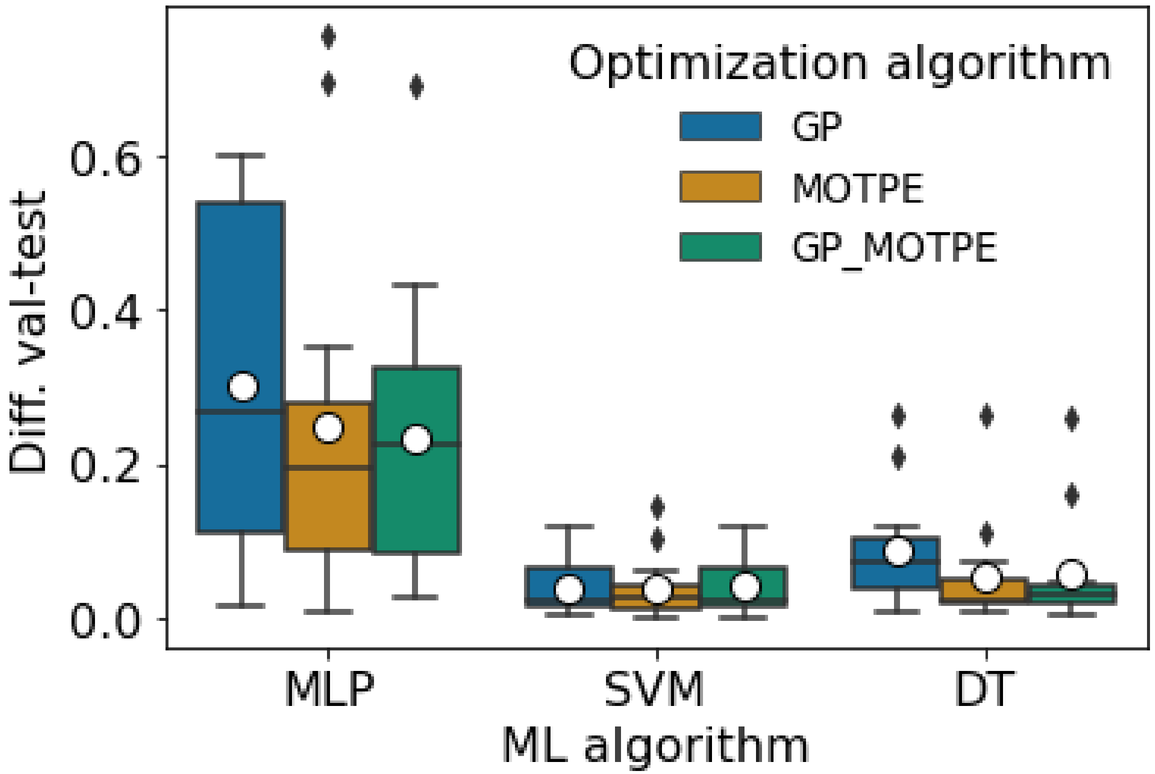




Results

Experiment 2: Hyperparameter optimization. Hypervolume and performance generalization analysis





- Our algorithm suggests a set of non-dominated HP configurations with the highest hypervolume in 15 trials
- HP configurations suggested by GPR are less reliable according to the difference between the validated and generalized hypervolume

Final remarks

- Hybrid algorithm that favour new HP configurations that are likely to be non-dominated, and that are expected to cause the maximum improvement in the scalarized objective function

- Our approach performed relatively well on (general) analytical test problems, yet the performance on the considered HPO problems varies amongst datasets and ML algorithms (*Not free lunch theorem*)

- GP_MOTPE showed promising reliability properties (small changes in hypervolume when the ML algorithm is evaluated on the test set)

Future works

- Handling uncertainty directly with TPE

- Analyse the performance of our algorithm with different sources of uncertainty and in more complex problems

Thanks

Q/A

Multi-objective hyperparameter optimization with performance uncertainty

Alejandro Morales-Hernández [1,2,*], Inneke Van Nieuwenhuyse [1,2], Gonzalo Nápoles [3]
[1] Core Lab VCCM, Flanders Make, Belgium
[2] Data Science Institute, Hasselt University, Belgium
[3] Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands

International Conference in Optimization and Learning, July 18th-20th, Syracuse, Italy

** Contact to: alejandro.moraleshernandez@uhasselt.be*