Surrogate Model-based Optimization in Practice

Thomas Bartz-Beielstein

SPOTSeven Lab
www.spotseven.de

Technology
Arts Sciences
TH Köln

Overview

Technology Arts Sciences TH Köln

Model-based optimization (MBO)

- Prominent role in todays modeling, simulation, and optimization processes
- Most efficient technique for expensive and time-demanding real-world optimization problems
- Engineering domain, MBO is an important practice

Example

- Waste heat boiler:
- CFD-optimized design

Model-based optimization (MBO)

- Recent advances in
  - computer science,
  - statistics, and
  - engineering
  - in combination with progress in high-performance computing
- Tools for handling problems, considered unsolvable only a few decades ago

Technology Arts Sciences TH Köln

Global optimization (GO)

- GO can be categorized based on different criteria.
- Properties of problems
  - continuous versus combinatorial
  - linear versus nonlinear
  - convex versus multimodal, etc.
- We present an algorithmic view, i.e., properties of algorithms
- The term GO will be used in this talk for algorithms that are trying to find and explore global optimal solutions with complex, multimodal objective functions [Preuss, 2015].
- GO problems are difficult: nearly no structural information (e.g., number of local extrema) available
- GO problems belong to the class of black-box functions, i.e., the analytic form is unknown
- Class of black-box function contains also functions that are easy to solve, e.g., convex functions

Problem

- *Optimization problem* given by

$$\text{Minimize: } f(\vec{x}) \quad \text{subject to } \vec{x}_l \leq \vec{x} \leq \vec{x}_u,$$

  where $f : \mathbb{R}^n \to \mathbb{R}$ is referred to as the *objective function* and $\vec{x}_l$ and $\vec{x}_u$ denote the lower and upper bounds of the search space (region of interest), respectively
- Setting arises in many real-world systems:
  - when the explicit form of the objective function f is not readily available,
  - e.g., user has no access to the source code of a simulator
- We cover *stochastic (random) search algorithms*, deterministic GO algorithms are not further discussed
- *Random* and *stochastic* used synonymously

Technology Arts Sciences TH Köln

Taxonomy of model-based approaches in GO

- [1] Deterministic
- [2] Random Search
  - [2.1] Instance based
  - [2.2] Model based optimization (MBO)
    - [2.2.1] Distribution based
    - [2.2.2] Surrogate Model Based Optimization (SBO)
      - [2.2.2.1] Single surrogate based
      - [2.2.2.2] Multi-fidelity based
      - [2.2.2.3] Evolutionary surrogate based
      - [2.2.2.4] Ensemble surrogate based

Overview

Random Search

- Stochastic search algorithm: Iterative search algorithm that uses a stochastic procedure to generate the next iterate
- Next iterate can be
  - a candidate solution to the GO or
  - a probabilistic model, where solutions can be drawn from
- Do not depend on any structural information of the objective function such as gradient information or convexity $\Rightarrow$ robust and easy to implement
- Stochastic search algorithms can further be categorized as
  - *instance-based* or
  - *model-based* algorithms [Zlochin et al., 2004]

[2.1] Instance-based Algorithms

- Instance-based algorithms: use a *single solution*, $\vec{x}$, or *population*, $P(t)$, of candidate solutions
- Construction of new candidates depends explicitly on previously generated solutions
- Examples: Simulated annealing, evolutionary algorithms

$t = 0$. InitPopulation(P).
Evaluate(P).
while not TerminationCriterion() **do**
    Generate new candidate solutions P'(t) according to a specified random mechanism.
    Update the current population P(t+1) based on population P(t) and candidate solutions in P'(t).
    Evaluate($P(t+1)$).
    $t = t + 1$.
end while

[2.2] MBO: Model-based Algorithms

- MBO algorithms: generate a population of new candidate solutions $P'(t)$ by sampling from a model
- In statistics: model $\equiv$ distribution
- Model (distribution) reflects structural properties of the underlying true function, say f
- Adapting the model (or the distribution), the search is directed into regions with improved solutions
- One of the key ideas: replacement of expensive, high fidelity, fine grained function evaluations, $f(\vec{x})$, with evaluations, $\hat{f}(\vec{x})$, of an adequate cheap, low fidelity, coarse grained model, M

Technology Arts Sciences TH Köln

[2.2.1] Distribution-based Approaches

- Metamodel is a distribution
- Generate a sequence of iterates (probability distributions) $\{p(t)\}$ with the hope that
$$p(t) \to p^* \text{ as } t \to \infty,$$
where p^*: limiting distribution, assigns most of its probability mass to the set of optimal solutions
- Probability distribution is propagated from one iteration to the next
- Instance-based algorithms propagate candidate solutions

1: $t = 0$. Let $p(t)$ be a probability distribution.
2: **while** not TerminationCriterion() **do**
3: Randomly generate a population of candidate solutions $P(t)$ from $p(t)$.
4: Evaluate($P(t)$).
5: Update the distribution using population (samples) $P(t)$ to generate a new distribution $p(t+1)$.
6: $t = t + 1$.
7: **end while**

[2.2.1] Estimation of distribution algorithms (EDA)

- EDA: very popular in the field of *evolutionary algorithms* (EA)
- Variation operators such as mutation and recombination replaced by a distribution based procedure:
  - Probability distribution estimated from promising candidate solutions from the current population ⇒ generate new population
- Larraaga and Lozano [2002] review different ways for using probabilistic models
- Hauschild and Pelikan [2011] discuss advantages and outline many of the different types of EDAs
- Hu et al. [2012] present recent approaches and a unified view

Technology
Arts Sciences
TH Köln

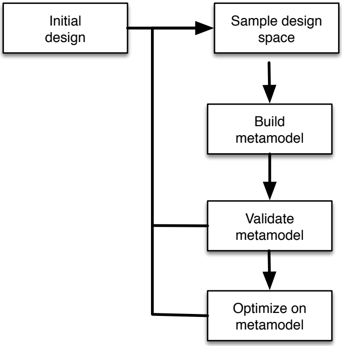
[2.2.2] Focus on Surrogates

- Although distribution-based approaches play an important role in GO, they will not be discussed further in this talk
- We will concentrate on surrogate model-based approaches
- Origin in statistical design and analysis of experiments, especially in response surface methodology [G E P Box, 1951, Montgomery, 2001]

[2.2.2] Surrogate Model-based Approaches

- In general: Surrogates used, when outcome of a process cannot be directly measured
- Imitate the behavior of the real model as closely as possible while being computationally cheaper to evaluate
- Surrogate models also known as
  - the cheap model, or
  - a response surface,
  - meta model,
  - approximation,
  - coarse grained model
- Simple surrogate models constructed using a data-driven approach
- Refined by integrating additional points or domain knowledge, e.g., constraints

[2.2.2] Surrogate Model-based Approaches



- Validation step (e.g., via CV) is optional
- Samples generated iteratively to improve the surrogate model accuracy

[2.2.2] Surrogate Model Based Optimization (SBO) Algorithm

$t = 0$. InitPopulation($P(t)$)
Evaluate($P(t)$)
while not TerminationCriterion() **do**
    Use $P(t)$ to build a cheap model $M(t)$
    $P'(t+1) =$ GlobalSearch($M(t)$)
    Evaluate($P'(t+1)$)
    $P(t+1) \subseteq P(t) + P'(t+1)$
    $t = t + 1$
end while

[2.2.2] Surrogates

- Wide range of surrogates developed in the last decades $\Rightarrow$ complex design decisions [Wang and Shan, 2007]:
  - (a) Metamodels
  - (b) Designs
  - (c) Model fit
- (a) Metamodels:
  - Classical regression models such as polynomial regression or response surface methodology [G E P Box, 1951, Montgomery, 2001]
  - support vector machines (SVM) [Vapnik, 1998],
  - neural networks [Zurada, 1992],
  - radial basis functions [Powell, 1987], or
  - Gaussian process (GP) models, *design and analysis of computer experiments*, Kriging [Schonlau, 1997], [Büche et al., 2005], [Antognini and Zagoraiou, 2010], [Kleijnen, 2009], [Santner et al., 2003]
- Comprehensive introduction to SBO in [Forrester et al., 2008]

Technology
Arts Sciences
TH Köln

[2.2.2] Surrogates: Popular metamodeling techniques

- (b) Designs [Wang and Shan, 2007]:

- Classical
  - Fractional factorial
  - Central composite
  - Box-Behnken
  - A-, D-optimal (alphabetically)
  - Plackett-Burmann
- Space filling
  - Simple grids
  - Latin hypercube
  - Orthogonal
  - Uniform
  - Minimax and Maximin
- Hybrid methods
- Random or human selection
- Sequential methods

[2.2.2] Surrogates: Popular metamodeling techniques

- (b) Designs: Sequential methods
- Model Refinement: Selection Criteria for Sample Points
- An initial model refined during the optimization $\Rightarrow$ Adaptive sampling
- Identify new points, so-called *infill points*
- Balance between
  - exploration, i.e., improving the model quality (related to the model, global), and
  - exploitation, i.e., improving the optimization and determining the optimum (related to the objective function, local)
- *Expected improvement* (EI): popular adaptive sampling method [Mockus et al., 1978], [Jones et al., 1998]

Technology Arts Sciences TH Köln

Model Refinement: Expected Improvement

Include Fig. 11 from Jones et al. [1998]

[2.2.2] Surrogates: Popular metamodeling techniques

- (c) Model fitting [Wang and Shan, 2007]:

- Weighted least squares regression
- Best linear unbiased predictor (BLUP)
- Likelihood
- Multipoint approximation
- Sequential metamodeling
- Neural networks: backpropagation
- Decision trees: entropy

Technology
Arts Sciences
TH Köln

[2.2.2] Applications of SBO

- Popular application areas: Simulation-based design of complex engineering problems
  - *computational fluid dynamics* (CFD)
  - *finite element modeling* (FEM) methods
- Exact solutions $\Rightarrow$ solvers require a large number of expensive computer simulations
- Two variants of SBO
  - (i) metamodel [2.2.2.1]: uses one or several different metamodels
  - (ii) multi-fidelity approximation [2.2.2.2]: same metamodel with different parameterizations

Technology Arts Sciences TH Köln

Example

- Automotive: Exhaust gas recirculation
- Optimization:
  - Pressure loss
  - Fill level: uniformly distributed
- 3D CFD tool to analyze behavior
- Surrogate substitutes CFD for optimization
- Evolutionary algorithm
- Project duration: several years

Technology Arts Sciences
TH Köln

[2.2.2.1] Applications of Metamodels and [2.2.2.2] Multi-fidelity Approximation

- Meta-modeling approaches
  - 31 variable helicopter rotor design [Booker et al., 1998]
  - Aerodynamic shape design problem [Giannakoglou, 2002]
  - Multi-objective optimal design of a liquid rocket injector [Queipo et al., 2005]
  - Airfoil shape optimization with CFD [Zhou et al., 2007]
  - Aerospace design [Forrester and Keane, 2009]
- Multi-fidelity Approximation
  - Several simulation models with different grid sizes in FEM [Huang et al., 2015]
  - Sheet metal forming process [Sun et al., 2011]
- "How far have we really come?" [Simpson et al., 2012]

[2.2.2.3] Surrogate-assisted Evolutionary Algorithms

- *Surrogate-assisted EA*: EA that decouple the evolutionary search and the direct evaluation of the objective function
- Cheap surrogate model replaces evaluations of expensive objective function

Example

- Electrostatic precipitator (filter)
- How to arrange baffles?
- Velocity profile

- Pressure loss and uniformity
- Velocities before and after electrostatic fields

Technology Arts Sciences TH Köln

Example

- Filter
- Velocity profile

- Pressure loss and uniformity
- Velocities before and after filter

[2.2.2.3] Surrogate-assisted Evolutionary Algorithms

- Combination of a genetic algorithm and neural networks for aerodynamic design optimization [Hajela and Lee, 1997]
- Approximate model of the fitness landscape using Kriging interpolation to accelerate the convergence of EAs [Ratle, 1998]
- *Evolution strategy* (ES) with neural network based fitness evaluations [Jin et al., 2000]
- Surrogate-assisted EA framework with online learning [Zhou et al., 2007]
- Not evaluate every candidate solution (individual), but to just estimate the objective function value of some of the neighboring individuals [Branke and Schmidt, 2005]
- Survey of surrogate-assisted EA approaches [Jin, 2003]
- SBO approaches for evolution strategies [Emmerich et al., 2002]

Technology Arts Sciences TH Köln

[2.2.2.4] Multiple Models

- Instead of using one surrogate model only, several models M_i, $i = 1, 2, \ldots, p$, generated and evaluated in parallel
- Each model $M_i : X \to y$ uses
  - same candidate solutions, X, from the population P and
  - same results, y, from expensive function evaluations
- Multiple models can also be used to partition the search space
  - The *tree-based Gaussian process* (TGP): regression trees to partition the search space, fit local GP surrogates in each region [Gramacy, 2007].
  - Tree-based partitioning of an aerodynamic design space, independent Kriging surfaces in each partition [Nelson et al., 2007]
- Combination of an evolutionary model selection (EMS) algorithm with expected improvement (EI) criterion: select best performing surrogate model type at each iteration of the EI algorithm [Couckuyt et al., 2011]

[2.2.2.4] Multiple Models: Ensembles

- Ensembles of surrogate models gained popularity:
- Adaptive weighted average model of the individual surrogates [Zerpa et al., 2005]
- Use the best surrogate model or a weighted average surrogate model instead [Goel et al., 2006]
- Weighted-sum approach for the selection of model ensembles [Sanchez et al., 2006]
  - Models for the ensemble chosen based on their performance
  - Weights are adaptive and inversely proportional to the local modeling errors

Technology Arts Sciences TH Köln

Overview

Technology Arts Sciences TH Köln

Model Selection Criteria

- EI approach handles the initialization and refinement of a surrogate model
- But not the selection of the model itself
- Popular *efficient global optimization* (EGO) algorithm uses a Kriging model
  - Because Kriging inherently determines the prediction variance (necessary for the EI criterion)
- But there is no proof that Kriging is the best choice
- Alternative surrogate models, e.g., neural networks, regression trees, support vector machines, or lasso and ridge regression may be better suited
- An *a priory* selection of the best suited surrogate model is conceptually impossible in the framework treated in this talk, because of the black-box setting

Technology Arts Sciences TH Köln

Single or Ensemble

- Regarding the model choice, the user can decide whether to use
  - one single model, i.e., one unique global model or
  - multiple models, i.e., an ensemble of different, possibly local, models

  The static SBO uses a single, global surrogate model, usually refined by *adaptive sampling*, but did not change ⇒ category [2.2.2.1]

Technology Arts Sciences TH Köln

Criteria for Selecting a Surrogate

- Here, we do not consider the selection of a new sample point (as done in EI)
- Instead: Criteria for the selection of one (or several) surrogate models
- Usually, surrogate models chosen according to their estimated true error [Jin et al., 2001], [Shi and Rasheed, 2010]
- Commonly used performance metrics:
  - *mean absolute error* (MAE)
  - root mean square error (RMSE)
- Generally, attaining a surrogate model that has minimal error is the desired feature
- Methods from statistics, statistical learning [Hastie, 2009], and machine learning [Murphy, 2012]:
  - Simple holdout
  - Cross-validation
  - Bootstrap

Overview

Criteria for Selecting a Surrogate: Evolvability

- Model error is not the only criterion for selecting surrogate models
- *Evolvability learning of surrogates* approach (EvoLS) [Le et al., 2013]:
  - Use fitness improvement for determining the quality of surrogate models
- EvoLS belongs to the category of *surrogate-assisted evolutionary algorithms* ([2.2.2.3])
- Distributed, local information

Technology
Arts Sciences
TH Köln

Evolvability Learning of Surrogates

- EvoLS: select a surrogate models that enhance search improvement in the context of optimization
- Process information about the
  - (i) different fitness landscapes,
  - (ii) state of the search, and
  - (iii) characteristics of the search algorithm to statistically determine the so-called *evolvability* of each surrogate model
- Evolvability of a surrogate model estimates the expected improvement of the objective function value that the new candidate solution has gained after a local search has been performed on the related surrogate model [Le et al., 2013]

Evolvability

- Local search: After recombination and mutation, a local search is performed
- It uses an individual local meta-model, M, for each offspring
- The local optimizer, φ_M, uses an offspring $\vec{y}$ as an input and returns $\vec{y}^*$ as the refined offspring
- Evolvability measure can be estimated as follows [Le et al., 2013]:

$$Ev_M(\vec{x}) = f(\vec{x}) - \sum_{i=1}^{K} f(\vec{y}_i^*) \times w_i(\vec{x})$$

with weights (selection probabilities of the offsprings):

$$w_i(\vec{x}) = \frac{P(\vec{y}_i \,|P(t), \vec{x})}{\sum_{j=1}^{K} P(\vec{y}_j \,|P(t), \vec{x})}$$

Technology
Arts Sciences
TH Köln

SPO

- EvoLS: distributed, local information. Now: more centralized, global information ⇒ *sequential parameter optimization* (SPO)
- Goal: Analysis and understanding of algorithms
- Early versions of the SPO [Bartz-Beielstein, 2003, Bartz-Beielstein et al., 2005] combined methods from
  - *design of experiments* (DOE) [Pukelsheim, 1993]
  - *response surface methodology* (RSM) [Box and Draper, 1987, Montgomery, 2001]
  - *design and analysis of computer experiments* (DACE) [Lophaven et al., 2002, Santner et al., 2003]
  - regression trees [Breiman et al., 1984]
- Also: SPO as an optimizer

Technology Arts Sciences TH Köln

SPO

- SPO: sequential, model based approach to optimization
- Nowadays: established parameter tuner and an optimization algorithm
- Extended in several ways:
  - For example, Hutter et al. [2013] benchmark an SPO derivative, the so-called *sequential model-based algorithm configuration* (SMAC) procedure, on the BBOB set of blackbox functions.
  - Small budget of $10 \times d$ evaluations of d-dimensional functions, SMAC in most cases outperforms the state-of- the-art blackbox optimizer CMA-ES

Technology
Arts Sciences
TH Köln

SPO

- The most recent version, SPO2, is currently under development
- Integration of state-of-the-art ensemble learners
- SPO2 ensemble engine:
  - Portfolio of surrogate models
  - regression trees and random forest, least angle regression (lars), and Kriging
  - Uses cross validation to select an improved model from the portfolio of candidate models
  - Creates a weighted combination of several surrogate models to build the improved model
  - Use stacked generalization to combine several level-0 models of different types with one level-1 model into an ensemble [Wolpert, 1992]
  - Level-1 training algorithm: simple linear model

Technology Arts Sciences TH Köln

SPO

- Promising preliminary results
- SPO2 ensemble engine can lead to significant performance improvements
- Rebolledo Coy et al. [2016] present a comparison of different data driven modeling methods
  - Bayesian model
  - Several linear regression models
  - Kriging model
  - Genetic programming
- Models build on industrial data for the development of a robust gas sensor
- Limited amount of samples and a high variance

Example: Sensor development

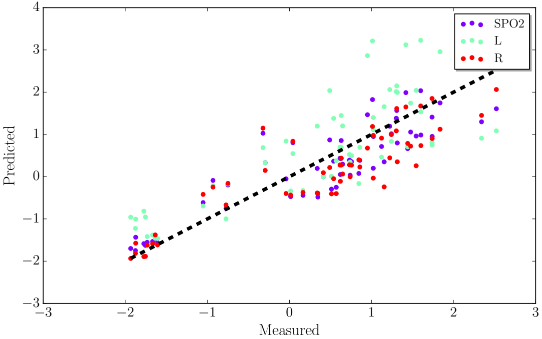
- Two sensors are compared
- 1st sensor (MSE)
  - Linear model (0.76), OLS (0.79), Lasso (0.56), Kriging (0.57), Bayes (0.79), and genetic programming (0.58)
  - SPO2 0.38
- 2nd sensor (MSE)
  - Linear model (0.67), OLS (0.80), Lasso (0.49), Kriging (0.49), Bayes (0.79), and genetic programming (0.27)
  - SPO2 0.28

Technology Arts Sciences TH Köln

Example: Sensor development

- Comparison of the mean squared error from the SPO2 ensemble and the single models:

```
SPO2  (MSE): 0.284948273406
L     (MSE): 0.673695001324
R     (MSE): 0.367652881967
```



Summary

- SMBO works!

Technology Arts Sciences TH Köln

Example

- Baffle geometry in electrostatic precipitators
- Combinatorial optimization problem: more than 2^{300} possible arrangements

Overview

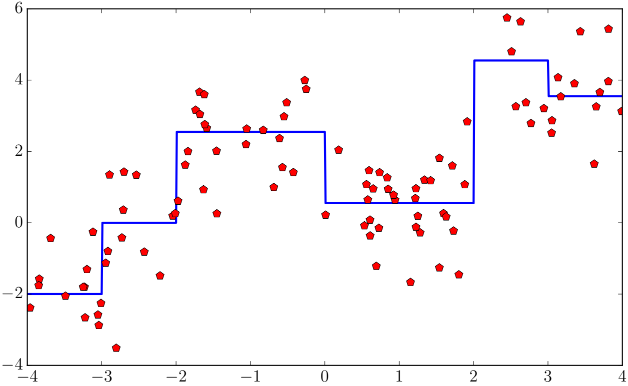
Function Definitions [jupyter]

- Motivated by van der Laan and Polley [2010], we consider six test functions
- All simulations involve a univariate X drawn from a uniform distribution in [-4, +4]
- Test functions:
  - f1(x): return -2 * I(x < -3) + 2.55 * I(x > -2) - 2 * I(x > 0) + 4 * I(x > 2) - 1 * I(x >3) + ϵ
  - f2(x): return 6 + 0.4 * x - 0.36x * x + 0.005x * x * x + ϵ
  - f3(x): return 2.83 * np.sin(math.pi/2 * x) + ϵ
  - f4(x): return 4.0 * np.sin(3 * math.pi * x) * I(x >= 0) + ϵ
  - f5(x): return x + ϵ
  - f6(x): return np.random.normal(0,1,len(x)) + ϵ
- $I(\cdot)$ indicator function, ϵ drawn from an independent standard normal distribution, sample size $r = 100$ (repeats)

Technology Arts Sciences TH Köln

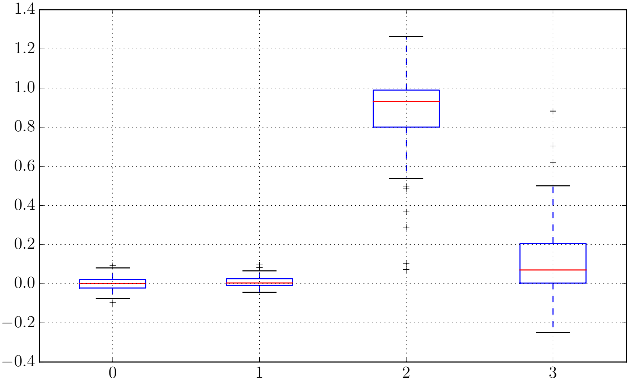
Function Definitions [jupyter]

- f1: Step function



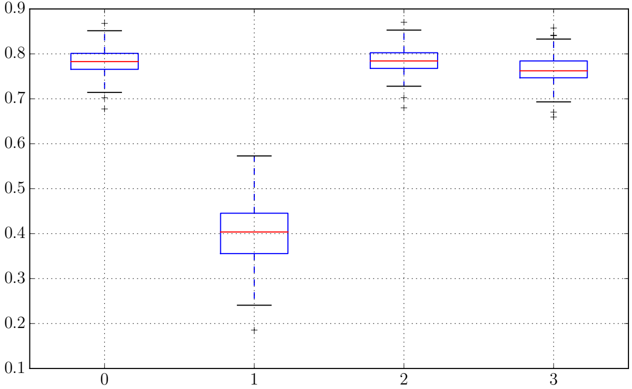
f1: Coefficients of the Level-1 Model [jupyter]

- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



f1: R^2 Values [jupyter]

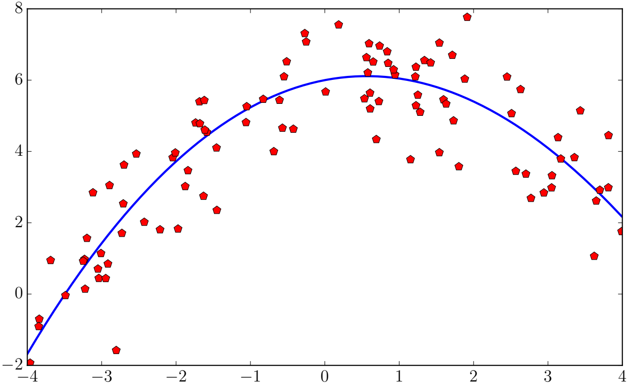
- R^2 (larger values are better) and standard deviation.
  - SPO: 0.78211976, 0.03308847
  - L: 0.4024831, 0.07134356
  - R: 0.78556947, 0.03187105
  - G: 0.76547433, 0.03564519



Technology Arts Sciences TH Köln

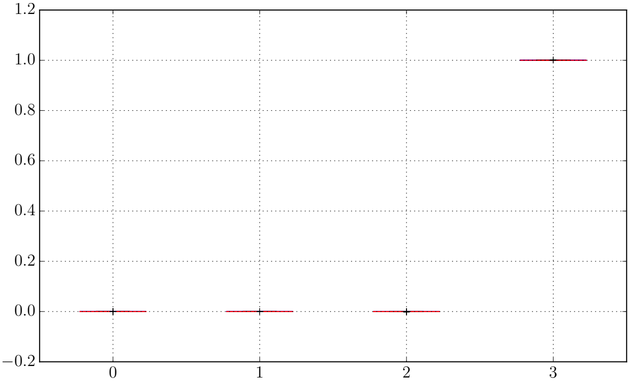
Function Definitions [jupyter]

- f2: Polynomial function



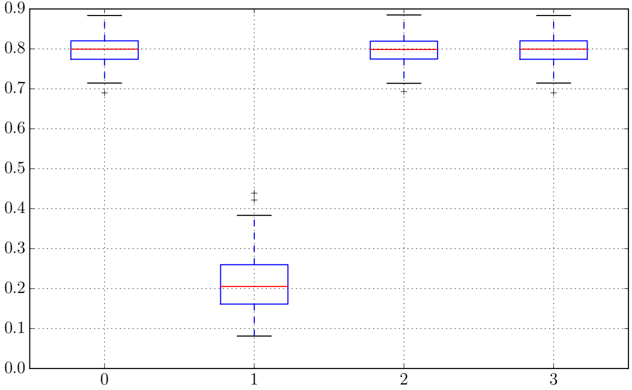
f2: Coefficients of the Level-1 Model [jupyter]

- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



f2: R^2 Values [jupyter]

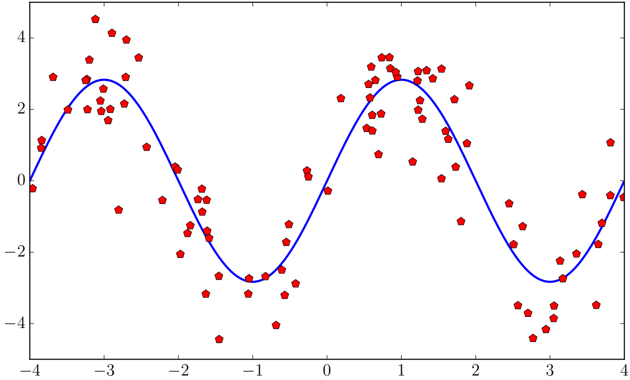
- R^2 (larger values are better) and standard deviation.
  - SPO: 0.79514735 0.03602018
  - L: 0.21445917 0.07656562
  - R: 0.79488344 0.03604606
  - G: 0.79514727 0.03602018



Technology Arts Sciences TH Köln

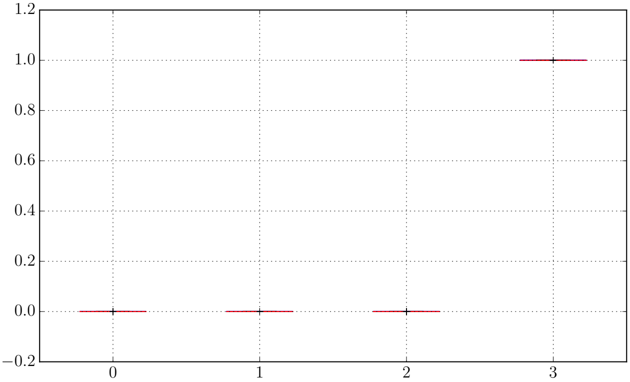
Function Definitions [jupyter]

- f3: Sine function



f3: Coefficients of the Level-1 Model [jupyter]

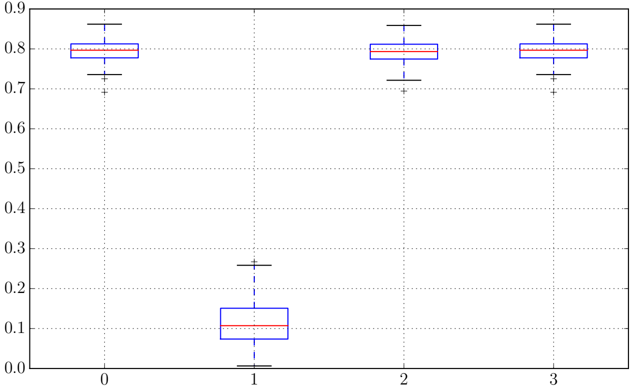
- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



Technology Arts Sciences TH Köln

f3: R^2 Values [jupyter]

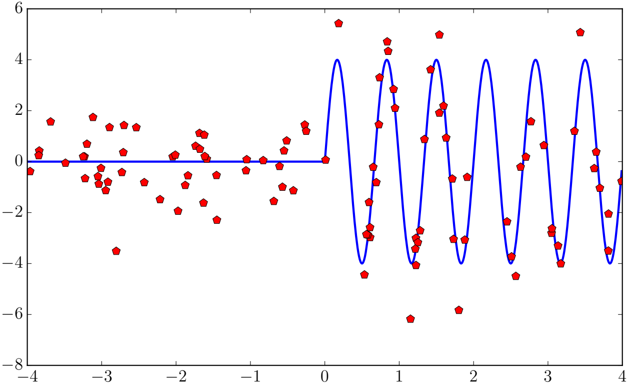
- R^2 (larger values are better) and standard deviation.
  - SPO: 0.7939634 0.02777211
  - L: 0.11677184 0.05688847
  - R: 0.79244941 0.02743085
  - G: 0.79396338 0.02777211



Technology Arts Sciences TH Köln

Function Definitions [jupyter]

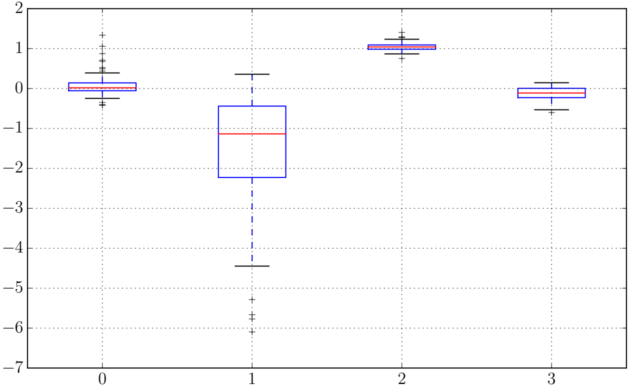
- f4: Composite function



Technology Arts Sciences TH Köln

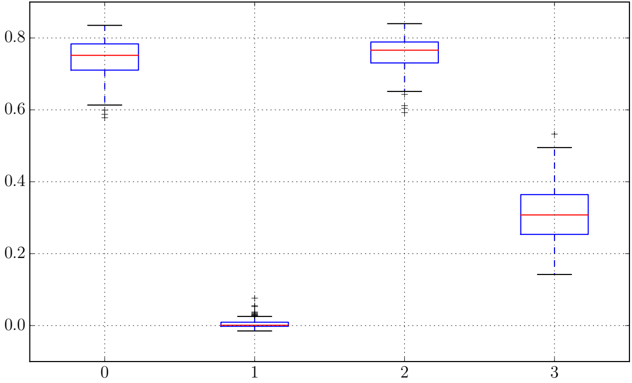
f4: Coefficients of the Level-1 Model [jupyter]

- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



f4: R^2 Values [jupyter]

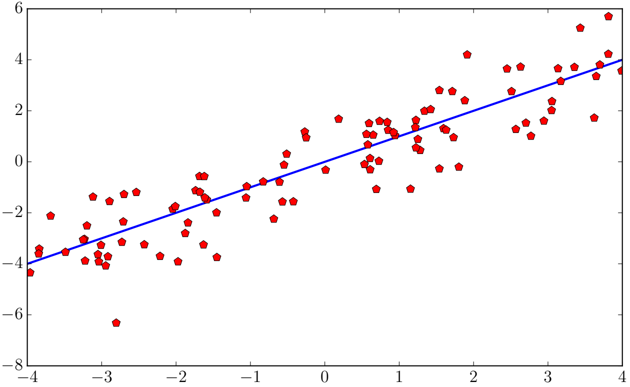
- R^2 (larger values are better) and standard deviation.
  - SPO: 0.74144195 0.05779718
  - L: 0.00651219 0.01489886
  - R: 0.75301025 0.05133169
  - G: 0.31721598 0.07939812



Technology Arts Sciences TH Köln

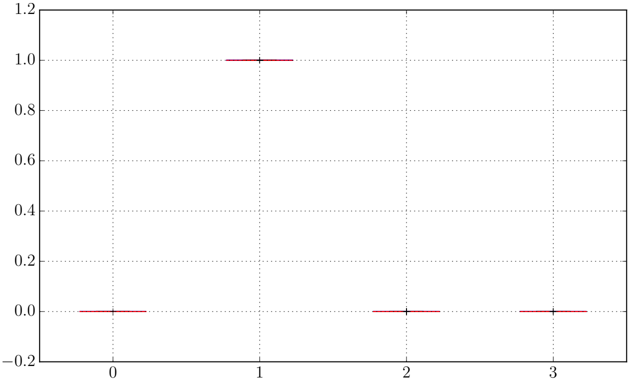
Function Definitions [jupyter]

- f5: Linear function



f5: Coefficients of the Level-1 Model [jupyter]

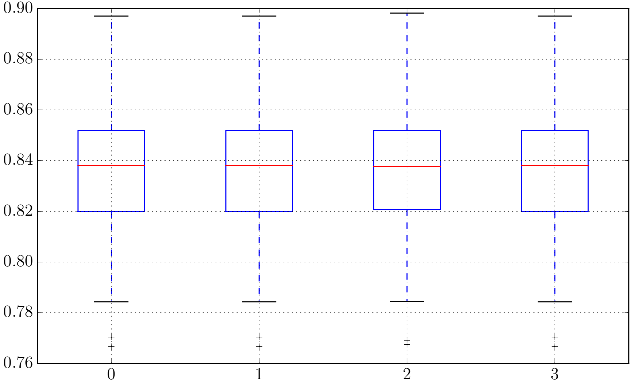
- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



Technology Arts Sciences TH Köln

f5: R^2 Values [jupyter]

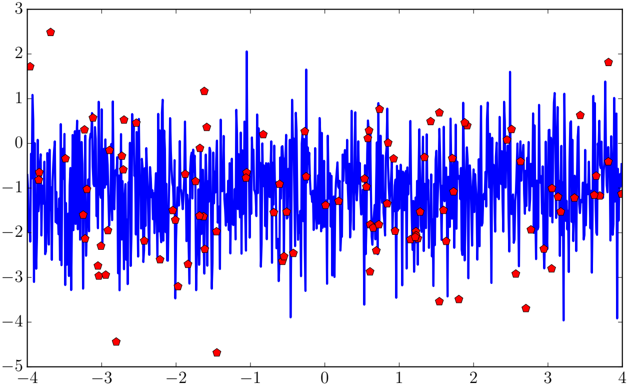
- R^2 (larger values are better) and standard deviation.
  - SPO: 0.8362937 0.02381472
  - L: 0.8362937 0.02381472
  - R: 0.83628043 0.02374492
  - G: 0.8362937 0.02381472



Technology Arts Sciences TH Köln

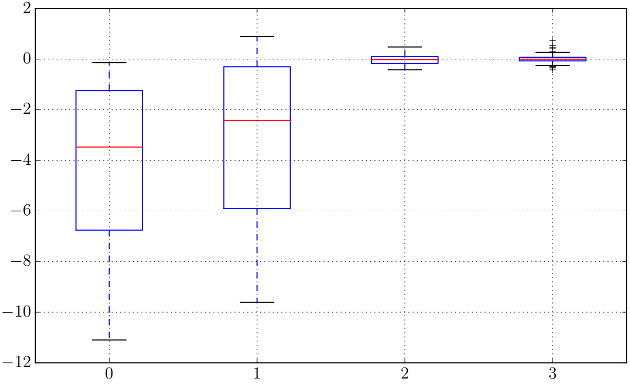
Function Definitions [jupyter]

- f6: Noise function



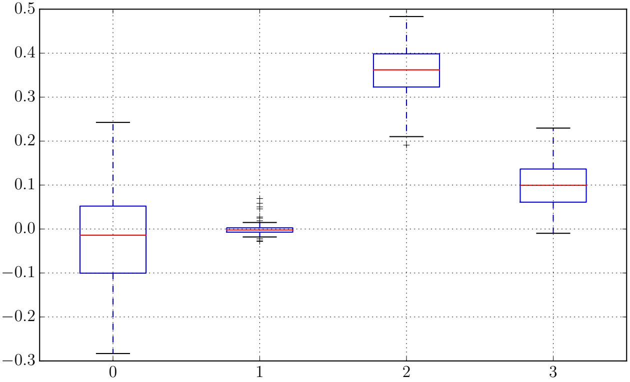
f5: Coefficients of the Level-1 Model [jupyter]

- The coefficients can be interpreted as weights in the linear combination of the models. 0 = intercept; 1, 2, and 3 denote the β_1, β_2, and β_3 values, respectively



f5: R^2 Values [jupyter]

- R^2 (larger values are better) and standard deviation.
  - SPO: -0.02025601 0.10308039
  - L: -0.00035958 0.01505964
  - R: 0.3586063 0.06232495
  - G: 0.10037904 0.05356867



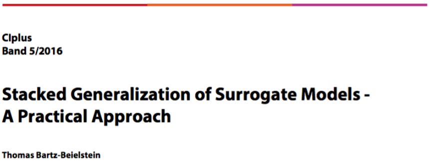
Overview

More: Video Lecture

- ▶ http://videolectures.net/bioma2016_bartz_beielstein_based_methods

Technology Arts Sciences
TH Köln

More: Publication



CIplus
Band 5/2016

Stacked Generalization of Surrogate Models - A Practical Approach

Thomas Bartz-Beielstein

- Bartz-Beielstein [2016], can be downloaded from http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hbz:832-cos4-3759

Acknowledgements

- This work has been supported by the *Bundesministeriums für Wirtschaft und Energie* under the grants KF3145101WM3 und KF3145103WM4.
- This work is part of a project that has received funding from the *European Union's Horizon 2020 research and innovation program* under grant agreement No 692286.

Technology Arts Sciences TH Köln

Alessandro Baldi Antognini and Maroussa Zagoraiou. Exact optimal designs for computer experiments via Kriging metamodelling. *Journal of Statistical Planning and Inference*, 140(9):2607–2617, September 2010.

Thomas Bartz-Beielstein. Experimental Analysis of Evolution Strategies—Overview and Comprehensive Introduction. Technical report, November 2003.

Thomas Bartz-Beielstein. Stacked Generalization of Surrogate Models—A Pratical Approach. Technical Report 05/2016, Cologne Open Science, Cologne, 2016. URL https://cos.bibl.th-koeln.de/solrsearch/index/search/searchtype/series/id/8.

Thomas Bartz-Beielstein, Christian Lasarczyk, and Mike Preuß. Sequential Parameter Optimization. In B McKay et al., editors, *Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland*, pages 773–780, Piscataway NJ, 2005. IEEE Press.

Andrew J Booker, J E Dennis Jr, Paul D Frank, David B Serafini, and Virginia Torczon. Optimization Using Surrogate Objectives on a Helicopter Test Example. In *Computational Methods for Optimal Design and Control*, pages 49–58. Birkhäuser Boston, Boston, MA, 1998.

G E P Box and N R Draper. *Empirical Model Building and Response Surfaces*. Wiley, New York NY, 1987.

J Branke and C Schmidt. Faster convergence by means of fitness estimation. *Soft Computing*, 9(1):13–20, January 2005.

L Breiman, J H Friedman, R A Olshen, and C J Stone. *Classification and Regression Trees*. Wadsworth, Monterey CA, 1984.

D Büche, N N Schraudolph, and P Koumoutsakos. Accelerating Evolutionary Algorithms With Gaussian Process Fitness Function Models. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 35(2):183–194, May 2005.

Ivo Couckuyt, Filip De Turck, Tom Dhaene, and Dirk Gorissen. Automatic surrogate model type selection during the optimization of expensive black-box problems. In *2011 Winter Simulation Conference - (WSC 2011)*, pages 4269–4279. IEEE, 2011.

Michael Emmerich, Alexios Giotis, Mutlu özdemir, Thomas Bäck, and Kyriakos Giannakoglou. Metamodel-assisted evolution strategies. In J J Merelo Guervós, P Adamidis, H G Beyer, J L Fernández-Villacañas, and H P Schwefel, editors, *Parallel Problem Solving from Nature—PPSN VII, Proceedings~Seventh International Conference, Granada*, pages 361–370, Berlin, Heidelberg, New York, 2002. Springer.

Alexander Forrester, András Sóbester, and Andy Keane. *Engineering Design via Surrogate Modelling*. Wiley, 2008.

Alexander I J Forrester and Andy J Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3): 50–79, January 2009.

K B Wilson G E P Box. On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(1):1–45, 1951.

K C Giannakoglou. Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence. *Progress in Aerospace Sciences*, 38(1):43–76, January 2002.

Tushar Goel, Raphael T Haftka, Wei Shyy, and Nestor V Queipo. Ensemble of surrogates. *Struct. Multidisc. Optim.*, 33(3):199–216, September 2006.

Robert B Gramacy. tgp: An R Package for Bayesian Nonstationary, Semiparametric Nonlinear Regression and Design by Treed Gaussian Process Models. *Journal of Statistical Software*, 19(9):1–46, June 2007.

P Hajela and E Lee. Topological optimization of rotorcraft subfloor structures for crashworthiness considerations. *Computers & Structures*, 64(1-4): 65–76, July 1997.

Trevor Hastie. *The elements of statistical learning : data mining, inference, and prediction*. Springer, New York, 2nd ed. edition, 2009.

Mark Hauschild and Martin Pelikan. An introduction and survey of estimation of distribution algorithms. *Swarm and Evolutionary Computation*, 1(3): 111–128, September 2011.

Jiaqiao Hu, Yongqiang Wang, Enlu Zhou, Michael C Fu, and Steven I Marcus. A Survey of Some Model-Based Methods for Global Optimization. In Daniel Hernández-Hernández and J Adolfo Minjárez-Sosa, editors, *Optimization, Control, and Applications of Stochastic Systems*, pages 157–179. Birkhäuser Boston, Boston, 2012.

Edward Huang, Jie Xu, Si Zhang, and Chun Hung Chen. Multi-fidelity Model Integration for Engineering Design. *Procedia Computer Science*, 44: 336–344, 2015.

Frank Hutter, Holger Hoos, and Kevin Leyton-Brown. An Evaluation of Sequential Model-based Optimization for Expensive Blackbox Functions. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, pages 1209–1216, New York, NY, USA, 2013. ACM.

R Jin, W Chen, and T W Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Struct. Multidisc. Optim.*, 23 (1):1–13, December 2001.

Y Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing*, 9(1):3–12, October 2003.

Y Jin, M Olhofer, and B Sendhoff. On Evolutionary Optimization with Approximate Fitness Functions. *GECCO*, 2000.

D R Jones, M Schonlau, and W J Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13: 455–492, 1998.

Jack P C Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3):707–716, February 2009.

P Larraaga and J A Lozano. *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*. Kluwer, Boston MA, 2002.

Minh Nghia Le, M N Le, Yew Soon Ong, Y S Ong, S Menzel, Stefan Menzel, Yaochu Jin, Y Jin, B Sendhoff, and Bernhard Sendhoff. Evolution by adapting surrogates. *Evolutionary Computation*, 21(2):313–340, 2013.

S N Lophaven, H B Nielsen, and J Søndergaard. DACE—A Matlab Kriging Toolbox. Technical report, 2002.

J Mockus, V Tiesis, and A Zilinskas. Bayesian Methods for Seeking the Extremum. In L C W Dixon and G P Szegö, editors, *Towards Global Optimization*, pages 117–129. Amsterdam, 1978.

D C Montgomery. *Design and Analysis of Experiments*. Wiley, New York NY, 5th edition, 2001.

K P Murphy. Machine learning: a probabilistic perspective, 2012.

Andrea Nelson, Juan Alonso, and Thomas Pulliam. Multi-Fidelity Aerodynamic Optimization Using Treed Meta-Models. In *Fluid Dynamics and Co-located Conferences*. American Institute of Aeronautics and Astronautics, Reston, Virigina, June 2007.

MJD Powell. Radial Basis Functions. *Algorithms for Approximation*, 1987.

Mike Preuss. *Multimodal Optimization by Means of Evolutionary Algorithms*. Natural Computing Series. Springer International Publishing, Cham, 2015.

F Pukelsheim. *Optimal Design of Experiments*. Wiley, New York NY, 1993.

Nestor V Queipo, Raphael T Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, January 2005.

Alain Ratle. Parallel Problem Solving from Nature — PPSN V: 5th International Conference Amsterdam, The Netherlands September 27–30, 1998 Proceedings. pages 87–96. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

Margarita Alejandra Rebolledo Coy, Sebastian Krey, Thomas Bartz-Beielstein, Oliver Flasch, Andreas Fischbach, and Jörg Stork. Modeling and

Optimization of a Robust Gas Sensor. Technical Report 03/2016, Cologne Open Science, Cologne, 2016.

E Sanchez, S Pintos, and N V Queipo. Toward an Optimal Ensemble of Kernel-based Approximations with Engineering Applications. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 2152–2158. IEEE, 2006.

T J Santner, B J Williams, and W I Notz. *The Design and Analysis of Computer Experiments*. Springer, Berlin, Heidelberg, New York, 2003.

M Schonlau. *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo, Ontario, Canada, 1997.

L Shi and K Rasheed. A Survey of Fitness Approximation Methods Applied in Evolutionary Algorithms. In *Computational Intelligence in Expensive Optimization Problems*, pages 3–28. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

Timothy Simpson, Vasilli Toropov, Vladimir Balabanov, and Felipe Viana. Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - Or Not. In *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, pages 1–22, Reston, Virigina, June 2012. American Institute of Aeronautics and Astronautics.

G Sun, G Li, S Zhou, W Xu, X Yang, and Q Li. Multi-fidelity optimization for sheet metal forming process. *Structural and Multidisciplinary . . .* , 2011.

M J van der Laan and E C Polley. *Super Learner in Prediction*. UC Berkeley Division of Biostatistics Working Paper . . . , 2010.

V N Vapnik. Statistical learning theory. Wiley, 1998.

G Gary Wang and S Shan. Review of Metamodeling Techniques in Support of Engineering Design Optimization. *Journal of Mechanical . . .* , 129(4): 370–380, 2007.

David H Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, January 1992.

Luis E Zerpa, Nestor V Queipo, Salvador Pintos, and Jean-Louis Salager. An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates. *Journal of Petroleum Science . . .* , 47(3-4):197–208, June 2005.

Z Zhou, Y S Ong, P B Nair, A J Keane, and K Y Lum. Combining Global and Local Surrogate Models to Accelerate Evolutionary Optimization. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 37(1):66–76, 2007.

Mark Zlochin, Mauro Birattari, Nicolas Meuleau, and Marco Dorigo. Model-Based Search for Combinatorial Optimization: A Critical Survey. *Annals of Operations Research*, 131(1-4):373–395, 2004.

J M Zurada. Analog implementation of neural networks. *IEEE Circuits and Devices Magazine*, 8(5):36–41, 1992.