Two Challenges in Surrogate-modeling: Merging Surrogate-models into Ensembles and Dealing with Structured or Combinatorial Search Spaces

Contributed Talk at
Surrogate-Assisted Multi-Criteria Optimization (SAMCO) Workshop
Lorentz Center, Leiden, NL

Martina Friese and Martin Zaefferer

03.03.2016

Technology Arts Sciences TH Köln

SPOT SEVEN

A Brief Introduction

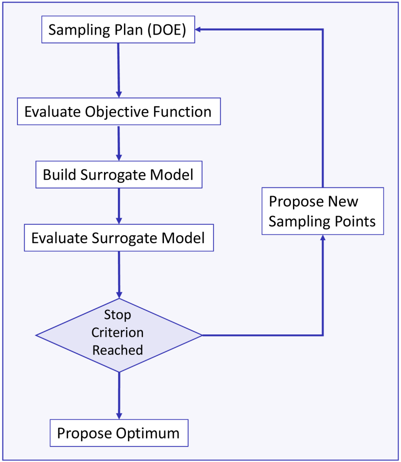


Abbildung: Sequential Optimization

- Surrogate-modeling is **state-of-the-art**
- **Large variety** of surrogate-models
- The **right choice** will make or break the parade

Research Questions

- Can linear combinations of predictors already improve as compared to model selection?
- Given the answer is positive, how can a system be build that finds the optimal linear combination on training data
- What are explanations of the observed behavior?

General Experiments Setup

- DACE Kriging Models using different kernel functions
  - Gaussian, exponential, spline
- Gaussian Landscape Generated objective functions

Tabelle: Gaussian landscape generator options used

Parameter	*Value*
Dimension	2 - 10
Number of peaks	10 - 40
Lower bounds	$\{0; 0\}$
Upper bounds	$\{5; 5\}$
Maximum function value	100
Ratio between global and local optima	0.8

Combining Two Surrogate-models

- two DACE Kriging models
  - (a) exponential kernel
  - (b) Gaussian kernel
- GLG objective function
  - 2-dimensional
  - 10 Gaussian components
- Latin Hypercube Design
  - size 40
- 50 Repetitions

- Linear combinations
  - $\alpha_n \in \{0.0, 0.1, 0.2, ..., 0.9, 1.0\}$
  - $\hat{y}_n = \alpha_n * \hat{y}_a + (1 - \alpha_n) * \hat{y}_b$
- Evaluation
  - Leave-one-out cross-validation
  - RMSE: $\sqrt{\frac{1}{n} \sum_{n=1}^{40} (y_n - \hat{y}_n)^2}$

First Results

- Base Models performed best in 6 out of 50 cases (a=4, b=2)
- RMSE values repetition-wise scaled for comparability

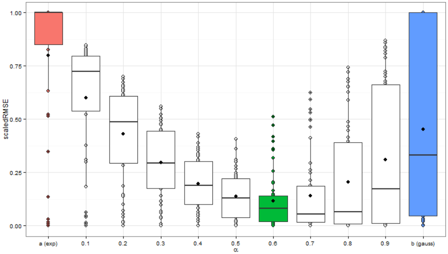
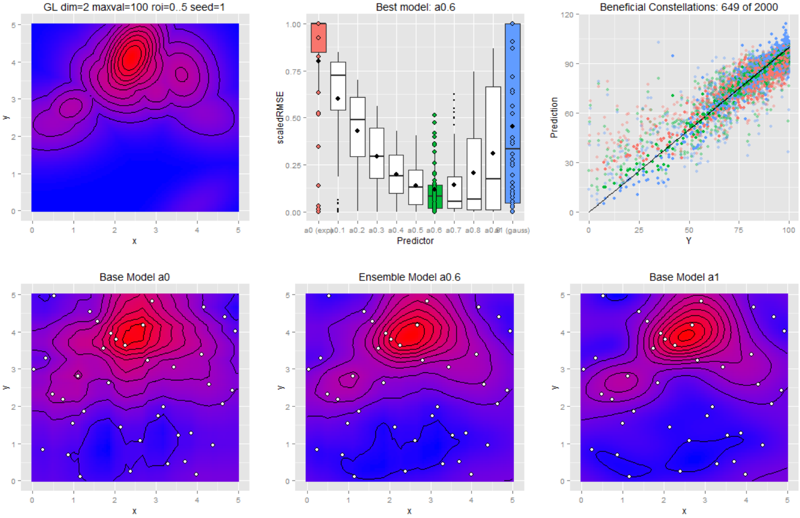


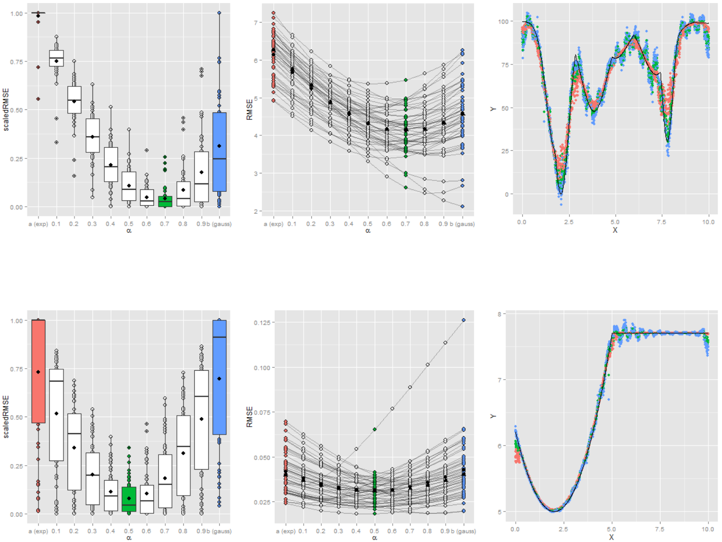
Abbildung: Performance results

- Best Model Choice: Ranking: Mean & Median & 3rd Quartile

2D Results

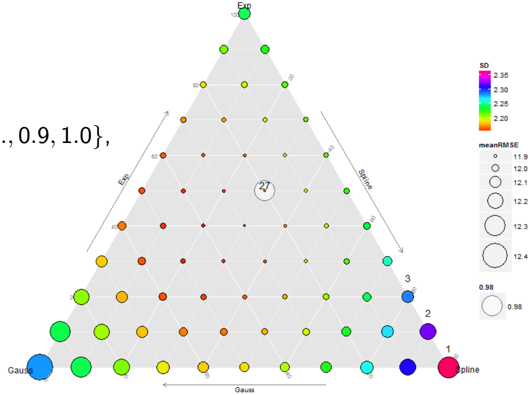

GL dim=2 maxval=100 roi=0..5 seed=1
Best model: a0.6
scaledRMSE
Predictor
a0 (exp) a0.1 a0.2 a0.3 a0.4 a0.5 a0.6 a0.7 a0.8 a0.9 a1 (gauss)
Beneficial Constellations: 649 of 2000
Prediction
Y
Base Model a0
Ensemble Model a0.6
Base Model a1
x
y

1D Results



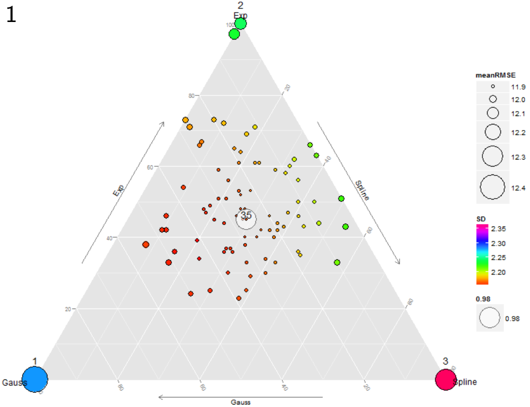
Combining three models

- two DACE Kriging models
  - exponential kernel
  - Gaussian kernel
  - spline kernel
- Linear combinations
  - $\alpha_n, \beta_n, \gamma_n \in \{0.0, 0.1, 0.2, ..., 0.9, 1.0\}$, $\alpha_n + \beta_n + \gamma_n = 1$
- GLG objective function
  - 4-dimensional
  - 40 Gaussian components



Heading for more

- Search space grows exponentially fast
- Linear combinations
  - $\alpha_n^1, \alpha_n^2, \ldots, \alpha_n^m \in \{0.0, 0.01, 0.02, ..., 0.99, 1.0\}$, where $\alpha_n^1 + \alpha_n^2 + \ldots, \alpha_n^m = 1$
- (1+1)-ES

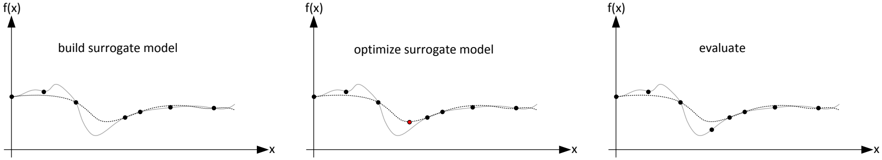


Outlook

- Continue this study for higher complexity cases
  - Number & Types of models
  - Dimensions of the objective functions
- Special interest on specific types of functions
- Implementation for sequential optimization processes

Motivation: Combinatorial Surrogate-models

- Well established in expensive, continuous optimization, e.g.,



What about expensive, combinatorial optimization problems?

- Example applications:
  - Engineering: weld path optimization [Voutchkov et al., 2005], twin-screw configuration [Teixeira et al., 2012]
  - Bioinformatics: protein sequence optimization [Romero et al., 2013]
  - Computer science: algorithm tuning/configuration [Hutter, 2009]

Types of Models

- Application specific models (expert knowledge, physics), e.g., [Voutchkov et al., 2005]
- Neural Networks
- Bayesian Networks
- Markov Random Fields [Allmendinger et al., 2015]
- Random Forest (Integer, Mixed Integer Problems) [Hutter, 2009]
- ...
- "Classical" kernel-based models:
  - Radial Basis Function Networks (RBFN) [Li et al., 2008; Moraglio and Kattan, 2011]
  - Support Vector Machines (SVM)
  - Kriging (Gaussian Processes) [Hutter, 2009; Zaefferer et al., 2014b]
- e.g., with Gaussian kernel and arbitrary distance:

$$k(x, x') = \exp(-\theta \mathrm{d}(x, x'))$$

Focus

- Kernel based methods, especially Kriging
- Powerful predictor
- Elegant parameter fitting (maximum likelihood estimation)
- Uncertainty estimate, Expected Improvement[1]

→ Efficient Global Optimization EGO [Jones et al., 1998]

[1] Other methods do also provide an uncertainty estimate, e.g., RBFN [Sóbester et al., 2005]

Does it work at all?

- Positive results [Zaefferer et al., 2014b,a]
- Genetic Algorithm (+Kriging)
- Inexpensive test-functions (permutation problems)
- Rather low-dimensional: $d \in [10, ..., 50]$

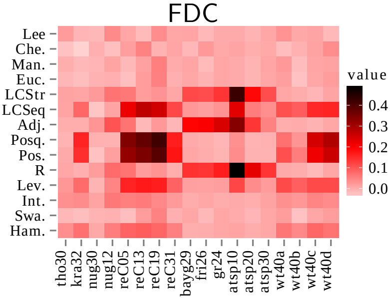
- Negative results [Pérez Cáceres et al., 2015]
- Ant Colony Optimization (+Kriging)
- Inexpensive test-functions (permutation problems)
- Rather high-dimensional: $d \in [50, ..., 100]$

d: number of elements in permutations

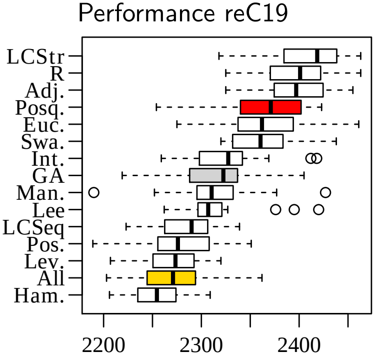
Choosing a distance

* Choice of distance measure crucial for success
* Use prior knowledge (if available?)
* Cross-validation

- Fitness Distance Correlation (FDC) (potentially misleading)



- MLE seems to work well (for Kriging)



All (MLE) - GA (model-free) - Posq (squared position distance)

Research Questions

- Which kernel/distance works best and why?
- How to choose a suitable kernel/distance?
- Or else, combine? (ensembles)
- Definiteness?
- Genotypic vs phenotypic distances? [Hildebrandt and Branke, 2014]
- Dimensionality issues? Dimensionality reduction?
- Comparison to other model types?
- Visualization?

Allmendinger, R., Coello, C. A. C., Emmerich, M. T. M., Hakanen, J., Jin, Y., and Rigoni, E. (2015). Surrogate-assisted multicriteria optimization (wg6). In Greco, S., Klamroth, K., Knowles, J. D., and Rudolph, G., editors, *Understanding Complexity in Multiobjective Optimization (Dagstuhl Seminar 15031) - Dagstuhl Reports*, volume 5, pages 96–163, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Hildebrandt, T. and Branke, J. (2014). On using surrogates with genetic programming. *Evolutionary Computation*, pages 1–25.

Hutter, F. (2009). *Automated configuration of algorithms for solving hard computational problems*. PhD thesis, University of British Columbia.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492.

Li, R., Emmerich, M. T. M., Eggermont, J., Bovenkamp, E. G. P., Bäck, T., Dijkstra, J., and Reiber, J. (2008). Metamodel-assisted mixed integer evolution strategies and their application to intravascular ultrasound image analysis. In *Congress on Evolutionary Computation (CEC'08), Proceedings*, pages 2764–2771, New York, NY, USA. IEEE.

Moraglio, A. and Kattan, A. (2011). Geometric generalisation of surrogate model based optimisation to combinatorial spaces. In *Proceedings of the 11th European Conference on Evolutionary Computation in Combinatorial Optimization*, EvoCOP'11, pages 142–154, Berlin, Heidelberg, Germany. Springer.

Pérez Cáceres, L., López-Ibáñez, M., and Stützle, T. (2015). Ant colony optimization on a limited budget of evaluations. *Swarm Intelligence*, pages 1–22.

Romero, P. A., Krause, A., and Arnold, F. H. (2013). Navigating the protein fitness landscape with gaussian processes. *Proceedings of the National Academy of Sciences*, 110(3):E193–E201.

Sóbester, A., Leary, S. J., and Keane, A. J. (2005). On the design of optimization strategies based on global response surface approximation models. *Journal of Global Optimization*, 33(1):31–59.

Teixeira, C., Covas, J. A., Stützle, T., and Gaspar-Cunha, A. (2012). Multi-objective ant colony optimization for the twin-screw configuration problem. *Engineering Optimization*, 44(3):351–371.

Voutchkov, I., Keane, A., Bhaskar, A., and Olsen, T. M. (2005). Weld sequence optimization: The use of surrogate models for solving sequential combinatorial problems. *Computer Methods in Applied Mechanics and Engineering*, 194(30-33):3535–3551.

Zaefferer, M., Stork, J., and Bartz-Beielstein, T. (2014a). Distance measures for permutations in combinatorial efficient global optimization. In Bartz-Beielstein, T., Branke, J., Filipic, B., and Smith, J., editors, *Parallel Problem Solving from Nature–PPSN XIII*, volume 8672 of *Lecture Notes in Computer Science*, pages 373–383, Cham, Switzerland. Springer.

Zaefferer, M., Stork, J., Friese, M., Fischbach, A., Naujoks, B., and Bartz-Beielstein, T. (2014b). Efficient global optimization for combinatorial problems. In *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation*, GECCO '14, pages 871–878, New York, NY, USA. ACM.