

SPOT: A Toolbox for Interactive and Automatic Tuning of Search Heuristics and Simulation Models in the R Environment

Thomas Bartz-Beielstein, Oliver Flasch, Patrick Koch, and Wolfgang Konen

Faculty of Computer Science and Engineering Science, Cologne University of Applied Sciences
{ thomas.bartz-beielstein | oliver.flasch | patrick.koch | wolfgang.konen}@fh-koeln.de

Sequential Parameter Optimization

Model-based investigations are common approaches in simulation and optimization. The *sequential parameter optimization* (SPOT) package¹ for R is a toolbox for tuning and understanding simulation and optimization algorithms (Bartz-Beielstein et al., 2005; Bartz-Beielstein, 2010b). R is a freely available language and environment for statistical computing (R Development Core Team, 2008).

SPOT was successfully applied to problems from theory and practice. It includes methods for interactive and automatic tuning. SPOT implements classical regression and analysis of variance techniques (Kleijnen, 1987, 2008); tree-based models such as CART and random forest (Breiman, 2001); Gaussian process models or Kriging (Sacks et al., 1989), and supports combinations of different meta-modeling approaches.

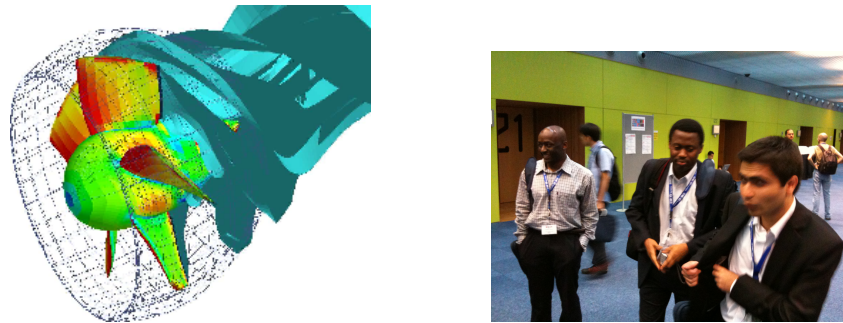


Figure 1: *Left*: Visualization of a linearjet propulsion system simulation. Due to symmetries, only a partial movement is simulated, and the water body is restricted to the parts in touch with the tube, rotor and stator. SPOT was used to optimize this model (Rudolph et al., 2009). *Right*: IEEE World Congress on Computational Intelligence in Barcelona 2010: Students from our research group after presenting their SPOT based results during one of the largest conferences in the field of computational intelligence. Three students received financial support from the International Office (Cologne University of Applied Science), Hans-Hermann-Voss-Stiftung, and from Prof. Bartz-Beielstein's research project FIWA, see http://www.f10.fh-koeln.de/Nachrichten/2010/07/f10_msg_05714.html

Strategic Decisions in the Tuning Process

SPOT uses a sequential tuning process, which can be divided into three major tasks:

1. Initial design
2. Meta modeling
3. Statistical analysis

¹The toolbox is available as an R package via CRAN <http://cran.r-project.org>

Each of these three tasks allows certain degrees of freedom, which will be exemplified in this article. Factorial (factorial) designs or space-filling designs can be used to accomplish the initial design task. Meta modeling can be based on several models, e.g., linear models, tree-based approaches (random forest), or Kriging. Finally, the statistical analysis can use simple tools from explorative data analysis, e.g., scatter plots, histograms, or modern tools from computational statistics.

SPOT is used by several research groups around the world. Prof. Konen and his team (research project SOMA) apply SPOT to develop a parameter-tuned data mining framework. Bartz-Beielstein (2010a) collects more than 100 publications related to the sequential parameter optimization.

Acknowledgements

This work is supported by the Bundesministerium für Bildung und Forschung (BMBF) under the grant FIWA (AIF FKZ 17N2309, "Ingenieurnachwuchs") and by the Cologne University of Applied Sciences under the research focus grant COSA.



References

- Bartz-Beielstein, T. (2010a). Sequential parameter optimization—an annotated bibliography. Technical Report 04/2010, Institute of Computer Science, Faculty of Computer Science and Engineering Science, Cologne University of Applied Sciences, Germany.
- Bartz-Beielstein, T. (2010b). SPOT: An R package for automatic and interactive tuning of optimization algorithms by sequential parameter optimization. Technical Report arXiv:1006.4645. CIOP TECHNICAL REPORT 05-10. COLOGNE UNIVERSITY OF APPLIED SCIENCES, Cologne University of Applied Sciences. Comments: Article can be downloaded from: <http://arxiv.org/abs/1006.4645>. Related software can be downloaded from <http://cran.r-project.org/web/packages/SPOT/index.html>.
- Bartz-Beielstein, T., Lasarczyk, C., and Preuß, M. (2005). Sequential parameter optimization. In McKay, B. et al., editors, *Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland*, volume 1, pages 773–780, Piscataway NJ. IEEE Press.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Kleijnen, J. P. C. (1987). *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York NY.
- Kleijnen, J. P. C. (2008). *Design and analysis of simulation experiments*. Springer, New York NY.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rudolph, G., Preuss, M., and Quadflieg, J. (2009). Double-layered surrogate modeling for tuning metaheuristics. In *ENBIS/EMSE Conference "Design and Analysis of Computer Experiments"*.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435.