Industrial applications of model-based simulation and optimisation

Thomas Bartz-Beielstein, Jörg Stork, Martin Zaefferer

17 May 2016, Ljubljana
Our Vision

• Scientific solutions to problems in industrial optimization and process analytics
• Individual task
• We are
  – passionate about technology, science, and critical reasoning
  – excited to promote modern data analysis and optimization technology into industry
  – proud that every success in our consulting projects brings a strong new relationship with industrial partners
• Our research motivates our students
Connections

• Close cooperation between industry and academia
• Well connected to leading research groups in the world
• Focused on solving engineering problems in
  – energy,
  – water,
  – steel,
  – automotive and
  – plastic industry
• We work independently, i.e., we do not sell one specific product or method
• Customers get the best from several worlds
How we work

• We do not claim to deliver a perfect solution in a single iteration, but measurable improvements in each step
• We rely on interaction and feedback
• Cutting edge research results
• Exploring new scientific frontiers
• Excellence in
  – teaching, data sciences, optimization, and statistical analysis are standard
Who we are

• Team of passionate data detectives thrilled by knowledge discovery
• Experience from different projects and form a unique team, combining complementary skills
• All different, have different backgrounds, have historically worked in different domains, but we all get a kick from having a problem solved
The SPOTSeven Process Model

- SPOTSeven defines a process model
- Extension and adaptation of the well-known SixSigma procedure
- SPOTSeven leaves room for individual procedures
- Enables transfer process (theory2practice)
1 Define

• Customer and business requirements specified from a high level point of view
• Current situation described
• Necessary steps defined
• Critical parameters and statistics identified
• Project team with clearly specified responsibilities
• Result: tentative project charter
• Developed during a workshop
2 Data Acquisition and Measurement

• Selection of suitable methods, instruments, and processes to collect data
• This step comprehends also a systematic gathering of measurement units, frequencies, and accuracies
• Important input-/output relations defined
• Based on current value: aimed value specified
3 Modeling and Analysis

• Functional relationships between in-/output parameters
• High-quality open source software such as R, as well as commercial products such as SAS/JMP, Minitab
• Process- and data analysis
  – Process analysis visualizes qualitative relationships, e.g., by integrating cause-effect diagrams
  – Data analysis is based on mathematical and statistical methods and relies on quantifiable measurement values
• Statistical tests
• In addition: interactive visualizations
4 Optimization

• Goal:
  – understandability
  – simplicity
  – interpretability
  – robustness

• Tools which allow an individual selection between exactness and robustness of the solution

• Communication with the practitioners
5 Integration and Deployment

• Implementation in real-world system.
• Verification of the process improvement on the real system
• New settings discussed with
  – project leaders,
  – technical experts, and
  – users to reach a high acceptance rate
6 Control

• Statistical quality control
• Documented solutions delivered
• Workshops and training courses:
  – discuss results and
  – to train practitioners
7 Meta Evaluation

• SPOTSeven itself contains a continuous improvement process
• Feedback from the customers
• Our passion:
  – to deliver the best results
  – learn from the process
  – teach students
Applying Design of Experiments for VOSS Automotive
VOSS Automotive

• Severe quality issues from an injection molding process
• Gasket ring
• Several million parts were delivered to automobile manufacturers every year
VOSS Automotive

- Design of Experiment (DoE)
- Here: small data
- Project duration: 20 weeks
- One student worked full time
- Weekly meetings:
  - project leaders, technical experts and users.
- Analysis and DoE based on Open Source Software
VOSS Automotive

Parameter

- Temperierung
- Heißkanalsystem
- Jahreszeit
- Zugluft
- Spätschicht
- Staudruck
- Frühschicht
- Einfahrausschuss
- Entlüftung
- Nachdruckumschaltzeitpunkt
- Einspritzgeschwindigkeit
- Temperatur
- Nachdruckhöhe
- Reinigung
- Nachtschicht
- Massetemperatur
- Kühlsystemtemperatur
- Werkzeugwandtemperatur
- Luftfeuchtigkeit
- Nachdruckzeit
- Schneekendrehzahl
VOSS Automotive

• Techniques:
  – Experimental design with two stages
  – Cause-effect diagrams
  – Screening
  – Optimization
  – Simple linear regression models and tree based models
  – Regression trees: easy to understand
  – Fractional factorial designs
  – Response surface methodology to visualize results

• Behavior of the system: explained with three parameters only
VOSS Automotive
Integration and Deployment:
- not the best configuration was selected to keep deterioration low
- Settings with reduced pressure and velocity were used instead
Experiments performed with settings found on the model
Control phase: confirmed that process stable
Development of a Multivariate Modelling and Online Adaptive Optimization for In-situ Measurements using an SO2-Sensor

Grant reference: KF3145101WM3

Total funding: 168,573,00 €

Funding: ”Zentrales Innovationsprogramm Mittelstand (ZIM)“-Kooperationen, Projektform: Kooperationsprojekt (KF)

Project Partner: ENOTEC GmbH

Goal:

• The Development of a multivariate modelling method for in-situ measurements.
• Freely parameterized and interpretable model. High robustness of the results
• Develop an optimization process which adapts the model continuously to the changing conditions in the measured gases.
IMProvT: Intelligente Messverfahren zur Prozessoptimierung
von Trinkwasserbereitstellung und Verteilung

Started Dec. 2015 until Dec. 2018
Total Funding: 590,445,00 €

Project Partners: TH Köln (GECO-C, SPOTSeven), DVGW-Technologiezentrum Wasser, Endress+Hauser, Thüringer Fernwasser, Landeswasserversorgung Stuttgart, Wasserversorgung Kleine Kinzig, IWW Zentrum Wasser, Aggerverband

Goal: development of solutions and tools for energy optimized drinking water production and distribution

• Optimization of sensor usage (positioning, reliability, drift detection)
• Application of machine learning methods for water quality monitoring
• Evaluation of energy saving potential for intelligent control systems
Surrogate-models for Combinatorial Search Spaces

Martin Zaefferer

17.05.2016
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., [Vouthkov 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:

$$k(x, x') = \exp(-\theta d(x, x'))$$
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]
Combinatorial Surrogate-models
First Results

Does it work at all?

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

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

\( d \): 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)

<table>
<thead>
<tr>
<th>Distance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ham.</td>
</tr>
<tr>
<td>Swa.</td>
</tr>
<tr>
<td>Int.</td>
</tr>
<tr>
<td>Lev.</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>Pos.</td>
</tr>
<tr>
<td>LCSeq</td>
</tr>
<tr>
<td>Adj.</td>
</tr>
<tr>
<td>Posq.</td>
</tr>
<tr>
<td>Adj.</td>
</tr>
<tr>
<td>LCStr</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>Man.</td>
</tr>
<tr>
<td>Euc.</td>
</tr>
<tr>
<td>LCStr</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>Ham.</td>
</tr>
</tbody>
</table>

- MLE seems to work well (for Kriging)

Performance reC19

- 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?


Thank you for your attention